DOCUMENT RESUME

ED 463 838 PS 030 238

AUTHOR	Barnow, Burt S.; Stapleton, David C.
TITLE	An Evaluability Assessment of Responsible Fatherhood Programs. Final Report.
INSTITUTION	Office of Child Support Enforcement (DHHS), Washington, DC.; Lewin Group, Inc., Washington, DC.; Johns Hopkins Univ., Baltimore, MD.
SPONS AGENCY	Office of the Assistant Secretary for Planning and Evaluation (DHHS), Washington, DC.; Ford Foundation, New York, NY.
PUB DATE	1997-06-06
NOTE	156p.; Project Officers were Linda Mellgren and Ronald Mincy.
AVAILABLE FROM	For full text: http://fatherhood.hhs.gov/evaluaby/intro.htm.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Cost Effectiveness; Data Collection; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Evaluation Problems; *Fathers; Formative Evaluation; Management Information Systems; *Parents; Program Descriptions; *Program Effectiveness; *Program Evaluation; Program Implementation; Sampling; Summative Evaluation
IDENTIFIERS	Evaluability Assessment

ABSTRACT

Noting the paucity of information on the effectiveness of the growing numbers of programs promoting responsible fatherhood, this report examines the evaluability of responsible fatherhood programs. The report provides policymakers with an evaluation design for use with a variety of responsible fatherhood programs, and gives direction to organizations supporting or conducting such evaluations. Activities informing the report included interviews with experts, a literature review on fatherhood issues, site visits to five programs, and input from experts on program evaluation. Chapter 1 of the report presents the aims of the fatherhood program, describes components of program evaluation, and discusses program readiness for evaluation. Chapter 2 describes procedures for conducting a process evaluation, including analysis and reporting. Chapter 3 discusses major design choices made in planning an impact evaluation. Chapter 4 describes potential outcomes of fatherhood interventions, suggests specific measures, and discusses difficulties that may be encountered while developing outcome measures. Chapter 5 includes a discussion of how and why explanatory variables are used in an impact analysis. Chapter 6 addresses issues related to selecting the sample and data collection methods. Chapter 7 discusses why a participation analysis should be conducted in conjunction with an impact evaluation and suggests appropriate methods. Chapter 8 discusses the analysis of the evaluation data that will be necessary to estimate the impacts of responsible fatherhood programs. Chapter 9 provides summary and concluding comments. The report's five appendices list the experts interviewed for the project, summarize site visits, provide sample discussion guide for conducting a process evaluation, contain preliminary evaluation findings from the Racine Goodwill Industries Fatherhood Program, and provide a technical discussion of participation and impact analysis methods. (KB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 838

An Evaluability Assessment of Responsible Fatherhood Programs: Final Report.

Burt S. Barnow and David C. Stapleton, Project Directors

Department of Health and Human Services,
Administration for Children and Families,
Office of Child Support Enforcement

August 6, 1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PS 030238

BEST COPY AVAILABLE

An Evaluability Assessment of Responsible Fatherhood Programs:

Final Report

By: THE LEWIN GROUP, INC. Project Directors: Burt S. Barnow, Johns Hopkins University and David C. Stapleton, The Lewin Group

Project Staff: Gina Livermore, The Lewin Group

Consultants:

- Jeffery Johnson, Management Plus
- John Trutko, James Bell Associates

Project Officers: Linda Mellgren for the Office of the Assistant Secretary for Planning and Evaluation, HHS and Ronald Mincy for The Ford Foundation.

August 6, 1997

CONTENTS

ACKNOWLEDGMENTS

CHAPTER ONE: INTRODUCTION AND BACKGROUND

I. Introduction

II. Fatherhood Programs and Evaluation Objectives

A. The Aim of Fatherhood Interventions

B. Why Evaluation is Important for Fatherhood Interventions

III. Components of Program Evaluation

A. Management Information Systems

B. Process Evaluation

C. Impact Evaluation

D. Cost-Benefit and Cost-Effectiveness Evaluations

IV. Program Readiness for Evaluation

A. Measurable Outcomes

B. Defined Service Components and a Hypothesized Relationship to Outcomes

C. Established Recruiting, Enrollment, and Participation Process

D. Understanding of the Characteristics of the Target Population, Program Participants, and Program Environment

E. Ability to Collect and Maintain Information

F. Adequate Program Size

V. Overview of the Remaining Chapters

CHAPTER TWO: PROCESS EVALUATION

I. Introduction

II. Purpose of a Process Evaluation

III. Questions Addressed by a Process Evaluation

IV. Methods for Collecting Information

A. Primary Data

B. Secondary Data

V. Analysis and Reporting

A. Analysis

B. Reporting

VI. Conclusion

CHAPTER THREE: MAJOR DESIGN ALTERNATIVES FOR AN IMPACT EVALUATION

I. Introduction

II. Experimental vs. Non-Experimental Designs

A. Experimental Design

B. Non-Experimental Design

C. Randomized Outreach Design

III. Single Site vs. Multiple Site Evaluations

A. Program Services

B. Program Objectives for Clients Served

C. Target Populations

IV. Summary

CHAPTER FOUR: OUTCOME MEASUREMENT

I. Introduction

II. Potential Outcomes and Methods of Measurement

A. Responsible Behavior

B. Father's Relationship With Child

C. Father's Support Capabilities

D. Child Well-Being

E. Co-parental or Team Relationship

III. Summary

CHAPTER FIVE: EXPLANATORY VARIABLES

I. Introduction

II. Purpose of Explanatory Variables

A. Estimator Precision

B. Controlling for Confounding Factors

C. Program Interactions with Individual Characteristics

D. Understanding the Determinants of Responsible Fatherhood Outcomes

III. Important Explanatory Variables and Their Measurement

A. Demographic Variables

B. Educational Attainment

C. Work History

D. Pre-Treatment Values of the Outcome Variables

E. Site-Specific Factors

F. Measures of Program Inputs

CHAPTER SIX: SAMPLING AND DATA COLLECTION

I. Introduction

II. Sampling Methodology

A. Defining the Target Population(s)

B. Recruiting Study Volunteers

C. Random Assignment to Treatment and Control Groups, and Program Enrollment

D. Sample Size

III. Methods of Data Collection

A. Surveys

B. Administrative Data

IV. Baseline and Follow-Up Data Collection

A. Types of Data to Collect

B. Timing of Data Collection

CHAPTER SEVEN: PARTICIPATION ANALYSIS

I. Introduction

II. Purposes of Participation Analysis

A. Increase Understanding of the Determinants of Participation

B. Control for Selection Effects

C. Assess the Effectiveness of Outreach/Recruiting Activities

III. Conducting Participation Analysis

A. Measuring Participation

B. Sample Descriptive Statistics

C. Multivariate Analysis of Participation

CHAPTER EIGHT: IMPACT ANALYSIS

I. Introduction

II. Methods for Analyzing Program Impacts at a Single Site

A. Analysis for a Continuous Outcome Variable under an Experimental Design

B. Application to a Non-Experimental Design

C. Modifications for a Randomized Outreach Design

D. Extension to Categorical and Limited Dependent Variables

III. Extension to a Multi-Site Impact Analysis

A. Experimental Design

B. Non-Experimental Design

C. Randomized Outreach Design

CHAPTER NINE: SUMMARY AND CONCLUSIONS

I. Introduction

II. Purpose of this Report

III. Features of Current Fatherhood Programs

A. Characteristics of Participants

B. Program Objectives

C. Service Delivery

D. Funding Streams

IV. Findings From An Evaluability Assessment of Selected Programs

A. Measurable Outcomes

B. Defined Service Components and a Hypothesized Relationship to Outcomes

C. Established Recruiting, Enrollment, and Participation Process

D. Understanding of the Characteristics of the Target Population, Program Participants, and Program Environment

E. Ability to Collect and Maintain Information

F. Adequate Program Size

V. Next Steps

APPENDIX A: LIST OF EXPERTS INTERVIEWED

I. Telephone Interviews

II. Participants at Practitioners Meeting

APPENDIX B: SITE VISIT SUMMARIES

The Cleveland Institute for Responsible Fatherhood and Family Revitalization

The Baltimore City Healthy Start Men's Services Program

The Baltimore St. Bernadine's Male Involvement Project

The Indianapolis Father Resource Program

The Racine Goodwill Industries Fatherhood Programs

APPENDIX C: PROCESS EVALUATION INTERVIEW GUIDES

Discussion Guide for Responsible Fatherhood Project Director and Sponsoring Organization Administrators

Discussion Guide for Responsible Fatherhood Program Managers and Staff

Discussion Guide for Community Human Service Providers

Discussion Guide for Organizations Providing Funding and Oversight for the Responsible Fatherhood Program

Discussion Guide for Responsible Father Program Participants

Discussion Guide for Community Leaders and Residents

APPENDIX D: EVALUATION OF THE RACINE GOODWILL INDUSTRIES FATHERHOOD PROGRAM

*Because of the mathematics symbols used in Appendix E, it is only available in .RTF format. Configure your browser to open an appropriate word processor if necessary.

APPENDIX E: TECHNICAL DESCRIPTION OF PARTICIPATION AND IMPACT ANALYSES

I. Participation Analysis

A. Participation Analysis under an Experimental Design

B. Participation Analysis in a Non-Experimental Design

C. Participation Analysis under a Randomized Outreach Design

D. Participation Analysis in a Multi-site Evaluation

II. Impact Analysis

A. A Model for a Continuous Outcome Variable under an Experimental Design

B. Application to a Non-Experimental Design

C. Modifications for a Randomized Outreach Design

D. Extension to Categorical and Limited Dependent Variables

E. An Econometric Model for Jointly Analyzing the Impacts of Multiple Programs

B. Non-Experimental Design

C. Randomized Outreach Design

ACKNOWLEDGMENTS

This project was sponsored by The Office of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services (DHHS) and The Ford Foundation. Linda Mellgren served as the Project Officer for ASPE and Ronald Mincy served as the Project Officer for Ford. They, and others at DHHS including Mark Fucello, Jeanine Smart, Barbara Cleveland, Anne Benson, and Barbara Cohen, provided valuable suggestions and guidance to the project.

This report benefitted greatly from the suggestions, based on a preliminary draft, received from the members of the project's Technical Review Group: Fred Doolittle (Manpower Demonstration and Research

Corporation), Ronald Ferguson (Kennedy School, Harvard University), and Jeffrey Smith (Department of Economics, University of Western Ontario).

We are also grateful for the generous assistance of the fatherhood program staff we interviewed during our site visits. We would especially like to thank: Charles Ballard, President of the National Institute for Responsible Fathers and Family Revitalization; Jerry Hamilton, Manager of Disadvantaged Programs, Goodwill Industries of Southeastern Wisconsin; Joseph Jones, Director of the Baltimore City Healthy Start Men's Services Program; Wallace McLaughlin, Director of the Wishard Memorial Hospital Father Resource Program in Indianapolis; and James Worthy, Director of the St. Bernadine's Head Start Male Involvement Project in Baltimore.

Finally, we would like to thank the many academic, government, and fatherhood program experts we interviewed who provided us with much useful information on parenting, child welfare, and fatherhood issues.

Top of Page

CHAPTER ONE

INTRODUCTION AND BACKGROUND

I. Introduction

The role of fathers has recently received increased attention from academics, the government, and private foundations. Fatherlessness is not only viewed as a cause of child poverty, but has also been shown to affect child development and children's prospects for academic and labor market success. There is also a perceived link between fatherlessness and social problems such as youth violence, domestic violence, and teen child bearing. The seriousness of father absence has prompted the federal government and organizations such as the Ford Foundation to begin funding programs that promote responsible fatherhood. There is, however, a paucity of evaluation information on the effectiveness of these programs.

The increased interest in programs that promote responsible fatherhood and the limited information currently available on the services provided and effectiveness of these programs has generated interest in the systematic evaluation of responsible fatherhood programs. For this reason, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services and the Ford Foundation have funded The Lewin Group and Johns Hopkins University to conduct an evaluability assessment of responsible fatherhood programs. The goal is to provide the Department and other policymakers with an evaluation design that can be used to evaluate a variety of responsible fatherhood programs. In addition, this report is intended to provide direction to organizations that would support or conduct evaluations by illustrating what is involved in the evaluation process and what mechanisms must be in place before a formal impact evaluation may be undertaken. It may also provide direction to programs that are building the capacity be evaluated.

In developing this report, we conducted several activities designed to learn more about fatherhood interventions and to identify the specific evaluation issues confronting these programs. These activities include:

- ***Interviews with Experts:*** We conducted phone interviews with nine experts on parenting, child welfare, and fatherhood issues. The experts include academics, program administrators, and policymakers. In addition, we convened a meeting with eight directors of fatherhood programs. The phone interviews and the meeting with the directors focused on identifying the goals of fatherhood programs, defining the key components of successful programs, and specifying important outcomes that should be assessed in an evaluation of responsible fatherhood programs. A list of the experts interviewed is in Appendix A.
- ***Review of Literature:*** We reviewed the literature on fatherhood issues with the primary purpose of identifying potential outcome measures that may be used in the evaluation of responsible fatherhood programs.
- ***Site Visits:*** We visited five fatherhood programs to obtain information from program staff, funders, and referring agencies on program goals and likely outcomes, characteristics of the intervention, characteristics of participants, and program administration. The programs we visited include: the Cleveland Institute for Responsible Fatherhood and Family Revitalization (IRFFR); the Baltimore Head Start Male Involvement Project (MIP); the Baltimore Healthy Start Men's Services Program (MSP); the Indianapolis Fathers Resource Program (FRP); and the Racine Goodwill Industries (RGI) fatherhood program. Appendix B contains a site visit summary for each program.
- ***Input from Technical Experts:*** We established a panel of three technical experts in the area of program evaluation who reviewed and commented on our preliminary evaluation design report.(1)

Their comments were received and discussed at a meeting convened at DHHS attended by the project officers, project staff, and a number of DHHS staff from several agencies.

In the remainder of this chapter, we provide a brief overview of the aim of fatherhood interventions; discuss the objectives of evaluating fatherhood programs; describe the major components of a program evaluation; and discuss some of the characteristics fatherhood programs must have in order to be ready for an evaluation. In the final section, we provide an overview of the remaining chapters of the report.

II. Fatherhood Programs and Evaluation Objectives

A. The Aim of Fatherhood Interventions

Many non-custodial fathers are responsible parents and want to be actively involved in the lives of their children. However, there may exist substantial barriers that prevent or inhibit a father's involvement with his child. The National Center on Fathers and Families identified seven core findings about fathers based on the experiences of front-line people working with fathers.(2) They include the following:

- Fathers care--even if caring is not always shown in conventional ways.
- Father presence matters--in terms of economic well being, social support, and child development.
- Joblessness is a major impediment to family formation and father involvement.
- Existing approaches to public benefits, child support enforcement, and paternity establishment operate to create obstacles and disincentives to father involvement. The disincentives are sufficiently compelling to have prompted the emergence of a phenomenon dubbed "underground fathers"--men who are involved in the lives of their children, but refuse to participate as fathers in formal systems.
- A growing number of young unwed fathers and mothers need additional support to develop the vital skills to share responsibility for parenting.
- The transition from biological father to committed parent has significant developmental implications for young fathers.
- The behaviors of young parents, both fathers and mothers, are significantly influenced by inter-generational beliefs and practices within families of origin.

The core findings provide an important context for understanding the unique challenges faced by young and adult men who want to become responsible fathers and the programs that help them achieve that goal.

In a recent publication, Jim Levine and Ed Pitt compiled the most extensive work to date on responsible fatherhood programs.(3) Their research and analysis of 300 community-based initiatives revealed characteristics common to the programs. Based on their findings, they offer the following strategic objectives as a framework for programs that promote responsible fatherhood:

- ***Prevent:*** Prevent men from having children before they are ready for the financial and emotional responsibilities of fatherhood.
- ***Prepare:*** Prepare men for the legal, financial, and emotional responsibilities of fatherhood.
- ***Establish:*** Promote paternity establishment at childbirth so that every father and child have, at a minimum, a legal connection.
- ***Involve:*** Reach out to men who are fathers, whether married or not, to foster their emotional connection to and financial support of their children.

- ***Support:*** Actively support fathers in the variety of their roles and in their connection with their children, regardless of their legal and financial status (married, unmarried, employed, and unemployed).

The Levine and Pitt framework provides a broad view of the aim of fatherhood interventions. Individual programs, however, vary substantially in both the specific outcomes they attempt to achieve and the activities they undertake to achieve them. Among the five programs we visited, we observed substantial variation in the numbers of fathers served, the recruiting methods used, the services fathers received, and program goals (see Appendix B). One common theme, however, was an underlying philosophy that in order to be an effective and responsible father, men needed first to develop the capacity to take care of themselves.

B. Why Evaluation is Important for Fatherhood Interventions

Fatherhood programs and emphasis on male parenting are relatively recent phenomena in the social service sector. Many of the programs currently in place are either very new or, if established, have been experimenting with new interventions or changing the program focus over time to meet the interests and objectives of funders. It is generally the case that fatherhood programs have not adequately documented their performance. This may be because of limited resources, a lack of experience with methods of measuring performance, or simply because the focus of program staff has been on serving fathers rather than proving that methods are effective. While program staff may believe that their activities are helping fathers and resulting in positive impacts on society, others, particularly funders, may be skeptical of evidence of program effectiveness that is limited to anecdotes.

Evaluations of responsible fatherhood programs can serve two important functions:

- provide information to outside agencies and organizations regarding the objectives and the effectiveness of their interventions, which may be used to attract and justify funding from these outside sources; and
- provide information to program staff that may be used to modify program design to more efficiently and effectively serve the fathers who use their services.

From the program funding perspective, the results of an evaluation can be used to attract and justify funding from outside sources. The results of an objective evaluation conducted using accepted scientific methods provide believable evidence of a program's effectiveness. In addition to using evaluation findings as evidence of effectiveness, programs can use the findings to demonstrate how their objectives are similar to the objectives of potential funders. Both of these are critical elements for convincing organizations that they should provide funding to a particular fatherhood intervention.

From the program design perspective, an evaluation can address a variety of questions, the answers to which can help program staff tailor their programs to more effectively serve their clients. Examples of questions that might be addressed through an evaluation include:

- What are the characteristics of the fathers served by the program?
- What are potential obstacles to participation?
- What are the impacts of the program on fathers and their children?
- What are the most effective methods for achieving desired outcomes?
- Does the program have any long-term effects on fathers and their families?

An evaluation design can provide a structured framework for collecting and analyzing the information necessary to answer these questions.

Systematic evaluation of fatherhood program outcomes is crucial to both program design and funding.

Conducting rigorous evaluations using standard scientific methods can assist program operators in effectively planning their programs to meet funding requirements, in improving their work with fathers, and in furthering the development of the field of fatherhood research and policy.

III. Components of Program Evaluation

There are three primary components to conducting a program evaluation: the process evaluation, the impact evaluation, and the cost-benefit or cost-effectiveness evaluation. In Chapter Two, we describe how and why a process evaluation should be conducted, and in Chapters Three through Eight we describe in great detail the steps necessary for conducting an impact evaluation. This report does not address cost-benefit or cost-effectiveness evaluations, but we include a brief discussion of them here because they represent the next logical step once process and impact evaluations have been conducted. In addition to these three components, an important element of the ability to conduct a program evaluation is having a management information system (MIS) in place that is capable of maintaining and processing some of the data necessary for an evaluation.

Before launching into the detailed discussion of process and impact evaluations in the subsequent chapters, we provide a brief overview of the primary evaluation components.

A. Management Information Systems

An automated system for tracking program participants is a precursor to any evaluation effort. A program management information system (MIS) is necessary to document a client's participation in the program, the services he receives and does not receive, and important outcomes related to program participation. If it cannot be shown from a cursory analysis of program administrative data that there are beneficial outcomes related to program participation, then there is often no point in conducting a full-scale impact evaluation. The ability to track a client's progress through the program, both in terms of the services he receives and changes in important outcomes, is not only necessary before an evaluation effort can be undertaken, but is also useful to program managers who may use the information to improve program effectiveness. Mature social service programs often have an MIS in place for administrative purposes, including quality control.

B. Process Evaluation

A process evaluation is the systematic collection and synthesis of information on the program environment and processes. It provides contextual information to support analyses of program outcomes, impacts, and costs. The types of information collected in a process evaluation are not only vital inputs for helping to assess program effects, but also provide feedback that can be helpful in efforts to refine the program intervention and to support replication of successful program components at other locations. A process evaluation can tell us if the underlying model for the program was implemented with integrity, as well as identify variations in treatment and participants. It can identify key similarities and differences across program sites in program objectives, participation levels, service delivery strategies, the environment, and a variety of other areas. A process evaluation can also suggest hypotheses to be tested in an impact evaluation.

The types of information collected under a process evaluation include information on: the social, economic, educational, and cultural environment in which the program operates; program goals and objectives; program strategies and interventions; major program components and services; clients' goals and objectives and their flow through the service delivery system; participant characteristics; and funding and referral sources. In general, the information collected for the process evaluation is more qualitative than quantitative in nature.

The steps involved in conducting a process evaluation include: determining the specific information to be collected/questions to be answered; identifying key program stakeholders; developing interview discussion guides; conducting interviews with program stakeholders; analyzing program administrative data; and reporting findings. The information and insights obtained through conducting a process evaluation are extremely useful to evaluators in designing and conducting an impact evaluation.

C. Impact Evaluation

An impact evaluation determines the extent to which a program causes change in the outcomes of interest. The concept of impact assessment implies that there are a set of defined objectives and criteria of success that may be used to measure the impact of the program. Impact evaluations are essential when there is an interest either in comparing different programs or in testing the effectiveness of new efforts to ameliorate a particular community problem.

To conduct an impact evaluation, the evaluator must develop a plan for collecting and analyzing data on program outcomes that will permit him or her to demonstrate that observed impacts are a function of the intervention and not a result of other factors. Impact analyses typically involve the comparison of outcomes for program participants to those of a comparison or control group. To undertake such a comparison, appropriate scientific methods and controls must be employed in the sampling, data collection, and data analysis steps to ensure that the estimated program impacts are unbiased.

Unless programs have a demonstrable impact, it is difficult to defend their implementation or continued operation. A rigorous impact evaluation provides information about the effectiveness of a particular program that may be used to modify and improve program design and to justify continued funding and operation.

The major steps involved in conducting an impact evaluation include the following:

- ***Determine the Measurable Outcomes of Interest:*** In designing and conducting an impact evaluation, the evaluator must first determine the primary program outcomes of interest. The outcomes chosen for the evaluation should be those that are most directly related to the program goals. They must also be defined such that they can be observed and quantified by the evaluator.
- ***Select Study Design/Determine Sample Size:*** Once the outcomes of interest have been identified, the evaluator must choose a study design. The type of design chosen (experimental, non-experimental, or some hybrid) will be a function of the program's recruiting and service characteristics, the number of persons served by the program, the program's target area, and the resources available to perform the evaluation. The sample size necessary to conduct the evaluation will primarily be a function of the program outcomes to be measured and the hypothesized impact of the program on those outcomes. The smaller the program impact, the greater the sample size necessary to discern the impact.
- ***Develop Data Collection Instruments:*** Information on program participants and comparison or control group members is collected through baseline and follow-up surveys. The survey instruments should capture information on the outcomes of interest, important demographic characteristics, and other variables related to the outcomes of interest. The instruments should be pre-tested to ensure that respondents understand and answer the questions in the manner intended.
- ***Establish Data Collection/Management Capability:*** An MIS must be in place to maintain electronic data files on the information collected for the evaluation. The program's MIS must be able to track individuals' participation in the program and should maintain ongoing information on the program outcomes of interest.

- ***Collect Data:*** Data collection for the evaluation will be both ongoing and episodic in nature. MIS data will continue to be collected while participants are in the program. Baseline surveys will be administered over time as new participants enroll in the program. Baseline surveys may be administered all at once or over time to control/comparison group members, depending on the study design. Follow-up surveys are administered to both program participants and control/comparison group members at some time interval after initial enrollment in the study. These surveys are usually administered by professional survey organizations, and do not require program staff.
- ***Analyze Data:*** Once the data has been collected, participation and impact analyses are conducted. Participation analysis compares the characteristics of program participants to those of eligible non-participants, individuals who drop out of the program, and control/comparison group members. Impact analysis compares outcomes for participants and non-participants using statistical methods to control for differences between the two groups.
- ***Report Findings***: The final step is to compile the results of the evaluation in a concise report that may be distributed to program managers, funders, and policymakers.

D. Cost-Benefit and Cost-Effectiveness Evaluations

Establishing the degree to which programs have an impact on desired outcomes, as is the purpose of an impact evaluation, is important to program managers, funders, and policymakers. What may be equally important is the comparison of program outcomes to their costs. A comparison of costs to benefits, whether done formally or informally, is inherent in decisions regarding whether to implement, expand, or continue any social program.

Cost-benefit and cost-effectiveness evaluations provide a formal framework for relating program costs to program outcomes. Cost-benefit evaluations address the issue of *economic efficiency*. In other words, what are the benefits (to individuals, funders, or society) of allocating resources to a particular program relative to the benefits of allocating those resources to any alternative endeavor. Cost-benefit evaluations attempt to translate all program benefits and costs into dollar values so that what is gained can be compared to what is be given up. A cost-benefit evaluation can answer questions such as:

- Do the total benefits of a fatherhood intervention exceed the total costs? And, if so,
- Are the net benefits at least as great as the net benefits that could be obtained from allocating the resources to any other program?

Cost-effectiveness evaluations are more limited in scope. They focus on the cost of producing a particular outcome. Here, the outcome or benefit need not be expressed in monetary values, as with a cost-benefit evaluation. Instead, the effectiveness of a program in attaining a particular outcome is related directly to the costs. Assuming that paternity establishment is the relevant outcome, a cost-effectiveness evaluation can answer questions such as:

- What is the cost of increasing the rate of paternity establishment by X% among fathers using fatherhood program services?;
- How does the cost of increasing the rate of paternity establishment by X% vary across types of fatherhood services provided?; or
- How does the cost of increasing the rate of paternity establishment by X% using fatherhood services compare to the cost of achieving the same goal through employment services?

In general, a cost-benefit analysis informs questions regarding whether or not an outcome should be pursued at all, while a cost-effectiveness analysis informs questions regarding the most effective method for achieving a desired outcome, assuming the decision to pursue that outcome has already been made.

Whether a cost-benefit evaluation, a cost-effectiveness evaluation, or both are conducted will depend on the specific questions a program, funder, or policymaker wants answered and the feasibility of conducting such evaluations. Cost-benefit evaluations are considerably more difficult to perform than cost-effectiveness evaluations because of the difficulty in putting dollar values on the benefits of social programs. Placing a dollar value on outcomes such as paternity establishment and improved father/child relationships is a difficult and controversial task. Cost-benefit analyses must often rely on strong assumptions made by the evaluator when benefits or costs cannot be easily determined. For this reason, cost-effectiveness evaluations are often a more feasible alternative. Neither cost-benefit nor cost-effectiveness evaluations should be undertaken, however, until program impacts have been quantified.

IV. Program Readiness for Evaluation

There are several important traits that programs must develop before a rigorous impact evaluation may be conducted. These include:

- Measurable outcomes;
- Defined service components and their hypothesized relationship to outcomes;
- An established recruiting, enrollment, and participation process;
- Understanding of the characteristics of the target population, program participants and program environment;
- Ability to collect and maintain information; and
- Adequate program size.

Below, we discuss why each of these is important to the evaluation process, and describe where the fatherhood programs we visited are in their development of each trait.

A. Measurable Outcomes

Fatherhood programs need to have clearly stated goals to guide the evaluation process. Program goals may be very broad or quite specific, but in either case, the evaluator must be able to translate the goals of the program into a set of measurable outcomes that can be analyzed in an evaluation of the program. The outcomes that are chosen will play a major role in determining the kinds of data that will be collected, the methods that will be used to collect that data, the required sample size, the methods used to conduct the analysis, and, hence, the cost and feasibility of conducting an evaluation.

Most of the fatherhood programs we visited were able to articulate a set of measurable outcomes believed to be influenced by the program. Among the most common were increased education and employment, reduced alcohol and drug use, improved parenting skills, and increased father involvement with his child(ren). Programs also cited some more difficult-to-measure outcomes, for example, improved attitudes or feelings toward children and improved social and family interactions.

One program had some difficulty defining a set of measurable outcomes influenced by program participation, mostly because the focus of the program was on general attitude change rather than on achieving more easily measured objectives. The primary goal of this program is to reconnect fathers with their children, or, in their words, "to turn the hearts of fathers to their children, and the hearts of children to their fathers." The underlying philosophy and secondary goal of the program is attitude change. Staff at this program believe that reconnecting fathers to their children will lead to changes in attitude and behavior leading to paternity establishment, job placement, and improved relationships with their child and the child's mother. For evaluation purposes, it is difficult or impossible to devise a measure of "turning hearts of fathers to their children" and vice versa. Attitude change is also difficult to measure, but consequences of attitude change, such as paternity establishment, employment, etc., can be measured. Staff were, however, somewhat hesitant to identify specific consequences that could be used in an evaluation of their program, although an assessment of potential program impacts had been previously conducted by outside researchers.

B. Defined Service Components and a Hypothesized Relationship to Outcomes

Before an evaluation is conducted, there should be an established, underlying model relating specific program services to specific outcomes. If a program cannot identify the mechanisms through which it affects outcomes, it may be that its services are not affecting the outcomes of greatest interest to the program. As discussed above, an impact evaluation should not be undertaken unless programs can demonstrate some beneficial change in outcomes among participants, and have a logical reason for attributing the change to program services.

In addition, if there is the intent to evaluate the effectiveness of specific service components, it is necessary to identify those components and characterize them in a manner that may be used to quantify their presence and impact on outcomes of interest. While this is not crucial to an evaluation of overall program outcomes, including information on service components can be useful in gaining a better understanding of the determinants of favorable program outcomes, and can be used to control for differences in treatments both within and across programs.

Of the programs we visited, all were able to define the services they offered and, with the exception of the one program described above, link those services to hypothesized impacts on a set of measurable outcomes. The specific services offered tend to change over time, however. All programs seemed to be in the process of adding new services or refining those already in place. This is probably because most of the programs we visited are only a few years old.

C. Established Recruiting, Enrollment, and Participation Process

Responsible fatherhood programs often recruit their participants through a variety of channels including the courts, welfare agencies, hospitals, mothers, media, and word of mouth. The method of recruitment is an important consideration in designing an evaluation as it can point to potential sources of selection bias, dictate the feasibility of an experimental evaluation approach, and offer innovative ways to derive a comparison group if a non-experimental approach is adopted. For these reasons, the recruiting methods must be thoroughly understood by the evaluator and must remain consistent throughout the evaluation process.

Determining when and how a father actually enrolls and begins participation in the program is also important in conducting an evaluation. There should be an identifiable event that marks the individual as a formal participant receiving the program treatment. If "partial" participants or non-participants are counted as full participants, the effects of the treatment may be underestimated in the evaluation. The enrollment process is also important to consider because it may be a source of selection bias. If programs are using criteria to select participants such that those allowed to participate are most likely to experience successful outcomes, then not controlling for this selection will lead to an overestimate of the program's effect.

Of the programs we visited, most have established recruiting and enrollment practices. Only one program is in the process of experimenting with new recruiting techniques, as it is having difficulty attracting participants. This program also has a rather lengthy pre-screening process that would be difficult to replicate in recruiting control group members if an evaluation were to be conducted. With respect to program participation, two of the programs we visited are having difficulty defining exactly who is an active participant in their program. This is because a number of men in their programs do not participate on a regular basis, periodically returning to the program after long intervals of non-participation.

D. Understanding of the Characteristics of the Target Population, Program Participants, and Program Environment

Having an understanding of the characteristics of the target population, the characteristics of program participants, and the economic, policy, and social environment in which the program operates is important in designing the evaluation. This information can assist the evaluator in developing the sampling methodology to ensure that a study sample representative of the target population is obtained. This information is also important in deciding which variables should be included in the data collection effort and subsequently used in the participation and impact analyses. Finally, an understanding of the characteristics of the population served and the program context can help evaluators interpret the findings once the evaluation has been conducted.

All of the programs we visited seemed to have a good understanding of the population they serve and the environment in which the program operates. Many of the program managers live in or near the neighborhoods in which they operate their programs. While all but one of the programs lack an MIS, most of the programs still produce descriptive statistics on important characteristics of their participants, such as age, race, education, marital status, employment, number of children, and paternity status. In addition, most of the program managers we met seemed to be very knowledgeable about and well-linked to other agencies in the community such as state and local health and welfare agencies, child support enforcement, the criminal justice system, and agencies providing specific services to persons with low income such as housing, employment services, legal services, medical care, and substance abuse treatment.

E. Ability to Collect and Maintain Information

As discussed above, a program MIS is necessary to document a client's participation in the program, the services he receives, and important outcomes related to program participation. The ability to track a client's progress through the program, both in terms of the services he receives and changes in important outcomes, is a necessity for conducting an evaluation.

Only one of the programs we visited has any kind of computerized tracking system, and its system was still being developed and modified at the time of our visit. Another program has an MIS, but it is being used only to track female clients enrolled in its primary program. No computerized tracking of male clients is currently conducted.

F. Adequate Program Size

In order to conduct an impact evaluation, there must be a sufficient number of individuals participating in the program to obtain a reasonable level of statistical precision when estimating the program impacts. The sample size necessary for conducting an evaluation will depend, in part, on the outcomes of interest. Outcomes with values that vary greatly among those in the study population will require a larger sample size for statistical precision. This is also true for program impacts that are small. The smaller the program impact, the greater the sample size necessary to detect it.

Most of the programs we visited serve a very small number of individuals, so it would be difficult for an evaluator to obtain statistically significant results. Only one program serves a relatively large number of fathers. The caseload of this program at the time of our visit was about 500 fathers. The program receives from 50 to 60 new referrals each month. This program is by far the exception. Three of the programs we visited serve only about 50 new fathers each year. In addition to simply serving more clients, there are ways to enhance sample size for evaluation purposes. If programs operate at multiple sites, or use a relatively homogeneous methods to serve fathers, then multiple sites may be pooled for the evaluation. Another way to increase sample size is to increase the period of recruiting study participants for the evaluation. There are some disadvantages (discussed in Chapter Six), however, to prolonged periods of recruiting in conducting an impact evaluation.

To summarize, most of the programs we visited appear not to be ready for a formal impact evaluation. This is due primarily to three factors: the programs are very new and still at the stage of refining recruiting methods and program services; the programs lack automated systems for tracking and reporting on clients; and the number of fathers served by most of the programs is very small.

V. Overview of the Remaining Chapters

The remainder of the report is organized as follows:

In Chapter Two, we describe the elements necessary for conducting a process evaluation. We begin with a brief overview of the reasons why conducting a process evaluation in conjunction with an impact evaluation is useful, and then describe the evaluation questions and major data sources that can and should be incorporated into a process evaluation of responsible fatherhood programs. We then provide a detailed description of various data collection methods that may be used for obtaining new and existing data. We also provide an overview of an automated participant-level data system that could be used by responsible fatherhood programs to track participant characteristics, service utilization, and outcomes. We conclude with examples of descriptive, comparative, and exploratory analyses that could be conducted to address key process evaluation questions.

In Chapter Three, we discuss two major design choices that must be made in the planning process for an impact evaluation. These choices concern: whether to use an experimental (i.e., randomized program assignment) or non-experimental design, or some hybrid; and whether to evaluate each individual site independently or to pool the data from multiple sites and evaluate them jointly. We describe the options and discuss criteria to be considered in making the choice between design alternatives. The main criteria we discuss include: feasibility, impact estimator bias, estimator precision, and cost. We conclude the chapter with a summary of the most important points with respect to these criteria for each design feature.

In Chapter Four, we describe potential outcomes of fatherhood interventions, suggest specific measures that may be used in an evaluation, and discuss difficulties that may be encountered when developing measures for outcomes of fatherhood interventions. In Chapter Five, we provide a similar discussion for explanatory variables, including a discussion of how and why explanatory variables are used in an impact analysis.

In Chapter Six, we address issues related to the selection of the study sample and methods for collecting data on study participants. We begin with a discussion of the process by which treatment and control/comparison groups may be selected and methods for determining sample size. We then describe methods available to evaluators for collecting data on study participants, including surveys and program administrative data sources. We conclude the chapter with a discussion of the content and timing of baseline and follow-up data collection efforts.

In Chapter Seven, we discuss reasons why a participation analysis should be conducted in conjunction with an impact evaluation of fatherhood interventions, and present methods that may be used to perform such analyses.

In Chapter Eight we discuss the analyses of the evaluation data that will be necessary to estimate the impacts of responsible fatherhood programs. We present methods of conducting analyses under each of the alternative evaluation designs. We also discuss methods for jointly analyzing the impacts of multiple programs.

In Chapter Nine we provide summary and concluding comments.

Finally, we include several Appendices to the report: In Appendix A, we list the experts interviewed for the project; Appendix B contains site visit summaries of the fatherhood programs we visited; In Appendix C, we provide sample discussion guides for conducting a process evaluation; Appendix D contains preliminary evaluation findings from the Racine Goodwill Industries program; and in Appendix E, we provide a technical discussion of the participation and impact analysis methods presented in Chapters Seven and Eight.

Return to ToC

---

1. The Technical Review Group members are: Fred Doolittle (Manpower Demonstration and Research Corporation), Ronald Ferguson (Kennedy School, Harvard University), and Jeffrey Smith (Department of Economics, University of Western Ontario).

2. See National Center on Fathers and Families (1994). "Fathers and Families: Building a Framework to Support Practice and Research," Concept Paper. Philadelphia, PA.

3. See Levine, Jim and Pitt, Ed (1995). New Expectations: Community Strategies for Responsible Fatherhood, Family and Work Institute. New York, NY.

CHAPTER TWO

PROCESS EVALUATION

I. Introduction

In this chapter we describe the elements necessary for conducting a process evaluation for a responsible fatherhood program. We begin with a brief overview of the reasons why conducting a process evaluation in conjunction with an impact evaluation is useful, and then describe the evaluation questions and major data sources that can and should be incorporated into a process evaluation of responsible fatherhood programs. We then provide a detailed description of various data collection methods that may be used for obtaining new (primary) and existing (secondary) data. We also provide an overview of an automated participant-level data system that could be used by responsible fatherhood programs to track participant characteristics, service utilization, and outcomes. The chapter concludes with examples of possible descriptive, comparative, and exploratory analyses that could be conducted to address key process evaluation questions.

II. Purpose of a Process Evaluation

A process evaluation provides contextual information to support analyses of program outcomes, net impacts, and costs. For example, it can provide information about how fathers are recruited to the program and how they are served once they are in the program. The types of information collected under a process evaluation are not only vital inputs for helping to assess program effects, but also provide feedback that can be helpful in efforts to refine the program intervention and to support replication of successful program components at other locations. A process evaluation can tell us if the underlying model for the program was implemented with integrity, as well as identify variations in treatment and participants. It can identify key similarities and differences across program sites in program objectives, participation levels, service delivery strategies, the environment, and a variety of other areas.

The major objectives of a process evaluation for a responsible fatherhood program should be to:

- describe the social, economic, educational, and cultural environment in which the program operates;
- identify program goals and objectives and the extent of variation in these objectives across sites;
- establish the underlying logic of the major program strategies and interventions (i.e., how the program interventions are expected to affect fathers involved in the program and their families);
- establish the sequence of events and other descriptive information about program design, development, and start-up;
- describe major program components/services (i.e., the program interventions as they actually operate within each site), including variances between what was originally planned and what actually occurred;
- capture participants' goals and objectives and how participants flow through the service delivery system, including how they may be referred for services outside the program;
- describe participant characteristics;
- describe client outcomes and changes from pre-participation outcomes;
- document costs; and
- document successful approaches and assess their replicability in other localities.

The information and insights obtained through conducting a process evaluation are extremely useful, and in many cases necessary, for evaluators to develop and conduct an impact evaluation.

III. Questions Addressed by a Process Evaluation

There are a number of questions that should be addressed by a process evaluation of responsible fatherhood programs. Each of these questions need to be addressed for each individual program being evaluated, and if there are multiple sites within a program being evaluated (e.g., IRFFR sites in Cleveland, San Diego, and other localities), then for each program site. Among the key questions that should be addressed by a process evaluation are the following:

- What are the overall objectives of the responsible fatherhood program?
- What external factors (i.e., social, educational, political, economic, and cultural) have affected the development and ongoing operations of the responsible fatherhood program?
- How have these external factors affected participation, outcomes, and costs of the program?
- What interventions (i.e., services, assistance) have program participants received, and how have these interventions been structured?
- How, why, and in what numbers do individuals participate in the responsible fatherhood program? What are the characteristics of those that do and do not participate, and how is "participation" in the program defined?
- What types of gross impacts (or outcomes) appear to result from the responsible fatherhood program interventions (i.e., what changes in outcomes occur for participants and their families from the time fathers enroll to the time they leave the program)?
- Is the program replicable in other communities and what, if any, strategies employed by the program should be replicated elsewhere?

In structuring a process evaluation of responsible fatherhood programs, specific evaluation questions could be broken down into the following categories: (1) program context, (2) program design and goals, (3) program implementation, (4) program components/services, (5) outreach, intake, and assessment, (6) client characteristics, (7) coordination/integration of services, (8) project staffing and staff development, (9) changes in outcomes, (10) program budget and costs, and (11) program replicability. Specific evaluation questions and potential data sources are displayed in ***Exhibit 2.1***. An "X" in the column opposite an evaluation question indicates that the source could provide data helpful in addressing the specific question under the process evaluation.

We present a very comprehensive set of questions. The effort required to answer them all is substantial, as will become evident in the following section. The evaluator may need to narrow the scope of the questions in order to focus the process evaluation and reduce costs. The process of providing more focus needs to be carried out early in the project and requires input from the program, the evaluator, funders, and other stakeholders in the evaluation.

IV. Methods for Collecting Information

A process evaluation of responsible fatherhood initiatives should include both primary data collection and use of existing data sources. Primary data should be collected by interviewing individuals knowledgeable about the program's design, start-up, and/or ongoing operations. These interviews should be supplemented by the collection of participant-level data (through, if possible, an automated client information system) and a systematic review of existing client files and program documents. The sections that follow address the overall strategies and methods that can be employed in collecting both primary and secondary data.

A. Primary Data

To develop an accurate, objective, and comprehensive understanding of each responsible fatherhood program being evaluated, it is recommended that, at a minimum, evaluators conduct interviews with the following groups:

- responsible fatherhood project director and relevant sponsoring organization administrators;
- responsible fatherhood program managers, staff, and consultants, including staff involved in outreach, assessment, ongoing case management, and direct provision of services;
- administrators/staff at agencies providing referrals to the responsible fatherhood program;
- administrators/staff at other linked human service agencies providing services for program participants (including child support enforcement, education, health, mental health, social services, vocational, and criminal justice agencies);
- current and past program participants (and, if possible, the children's mothers or family members), as well as fathers eligible for services who have not participated in the program; and
- community leaders and residents within the area served by the program.

It is important to not only interview responsible fatherhood program administrators and staff who are currently with the program, but also individuals who may no longer be part of the program, but can provide insights on initial design and start-up of the program and a reference point for how the program may have changed over the years since its inception.

In the following sections, we provide a brief description of the types of information each of these groups is best suited to provide.

1. Responsible Fatherhood Project Director and Sponsoring Organization's Administrators

During our visits to some of the programs, we observed substantial cross-site differences in underlying program strategies and services. These differences stem from a number of factors, including: the basic philosophies of the organization's sponsoring the initiatives; the size and geographic distribution of the populations served; the funding streams and goals of funding organizations; local resources and the economic an policy environment; and a host of other factors. Sponsoring organizations' philosophies had a considerable effect on the design and day-to-day operations of programs that we visited.

The goals of two programs offer a contrasting example. One program's primary goal is to reconnect fathers with their children. Underlying this basic philosophy is the strong belief that reconnecting fathers to their children will lead to changes in attitude and behavior leading to paternity establishment, job placement, and improved relations with their children and the children's mother. The program's philosophy embraces the view that a father has the inner capacity to solve his own problems -- and, therefore, the role of staff is to assist him through the process of self-discovery. In contrast, a second program's primary goals are: to develop the capacity of young fathers to become responsible and involved parents, wage-earners, and providers of child support; and to assist fathers with developing the skills and behaviors necessary to cooperate in the care of their children, regardless of the character of the relationship with the mother. There is a strong emphasis on building the skills necessary for the father to be able to financially support his child. A primary goal of the program is to place fathers in jobs upon completion of the program's six-week curriculum.

Each program (and site) also is likely to draw upon staff and resources available through its parent organization (e.g., sites may use forms, curriculum, and information systems developed by the sponsoring organization). Hence, it will be important to interview the organization's executive director and/or administrator responsible for oversight of the responsible fatherhood program site. The discussion guide found in Appendix C provides a series of questions that will help to structure this interview.

The sponsoring organization's executive director (and/or other administrator) is likely to be knowledgeable about the history of the funding for the program -- why the organization submitted a proposal for a specific site, what was initially intended in the program's design, and (perhaps) reasons why the sponsoring organization was selected. He or she should be able to explain how the responsible fatherhood initiative fits into the overall organization mission and how this mission guides the responsible fatherhood strategies and specific services or activities. The executive director may be able to provide a chronology of the program start-up (if he or she was with the organization at the time the program started), including identification of barriers encountered during the project start-up (e.g., possible resistance within the community or from other human service agencies) and how these barriers may have been overcome. Finally, the executive director is likely to have an understanding of the program's budget and how funds are allocated to major program components.

The site's project director (i.e., the individual at the site responsible for day-to-day oversight and direction) is likely to have the most comprehensive knowledge of operations at the site. The project director should be able to describe virtually all aspects of the site's operations, including outreach and intake, case management, client flow, the structure of major program components/services, linkages with other service providers, and types of fathers and families served by the program. He or she is likely to have views on ways in which the program has or has not been effective in serving the target population. If the project director has been with the program since its inception, he or she should be able to identify barriers to implementation and ways in which these barriers were overcome. Finally, the project director will be able to identify other individuals who should be interviewed as part of the process evaluation.

2. Responsible Fatherhood Program Managers and Staff

Responsible fatherhood program staff (e.g., intake workers, case managers, counselors, group leaders, MIS specialists, and clerical staff) can provide further details about major program components and services, as well as impressions about how the specific program interventions appear to be affecting fathers and their families. For example, because of their daily interaction with fathers, staff probably have views about which fathers have (and have not) been participating in the program and why, what are the most common client needs, and which of these needs the program is (and is not) addressing. The staff will be able to provide details about the specific services they are delivering (e.g., needs assessment, individual and/or family counseling, job placement, education, legal services, and parenting skills) and may have views on whether and to what extent specific services have affected fathers and their families. They will also be able to provide details on the process by which participants are matched to particular services. Some staff, particularly those working directly with fathers, will be able to provide contextual information about the families served and the surrounding community. Appendix C provides a discussion guide that will be helpful in structuring discussions with program managers and staff. During these discussions, it is important to tailor questions to the specific program components or services in which staff have been involved.

3. Community Human Service Providers

Other community human service providers refers to private or public agencies providing services within the community that the responsible fatherhood program is operating and that are needed and/or utilized by participants or their families. These services include child support enforcement, education, health and mental health services, vocational training, legal services, and a wide range of other social services. Some may have been providing these services prior to the project's inception and others may be new to the community and only recently linked to the fatherhood program. There is also the possibility that there are other providers of responsible fatherhood services in the same community. The discussion guide included in Appendix C can be helpful in structuring interviews with officials at these other service providers, though this instrument will need to be tailored to each specific interviewee and according to the types of services being provided by the linked agency.

Other service providers are an important source of information about available services in the community. If these service providers work directly with the responsible fatherhood program and receive frequent referrals of fathers or members of their families from the program, they will have specific information about the fathers' needs and their willingness to follow-up on referrals for services. The providers may be able to offer opinions about the quality and comprehensives of the responsible fatherhood program's services, as well as views on strategies or interventions that appear most effective in reducing risk factors for fathers. Finally, other providers may be able to provide indications of how well responsible fatherhood program services have been integrated into the fabric of services at the community level.

4. Organizations Providing Funding and Oversight for the Responsible Fatherhood Program

As nonprofit human service agencies, the organizations operating responsible fatherhood programs are likely to have received funding through one or more other organizations, such as state and local government agencies, the United Way, or other non-profit organizations. These organizations are likely to have played -- to varying degrees -- roles in the development, implementation, and ongoing operation of the program. For example, in addition to funding, they may have some (even considerable) input on the program objectives, eligibility rules, definition of the target area for participants, overall program design and types of services provided. In addition, these funding organizations may provide technical assistance, training, and ongoing program monitoring.

Administrators and staff of the funding agencies should be able to provide a chronology of program development, including original program goals, how sponsoring agencies and sites were selected, and an overview of program start-up at each site. Staff at these agencies may also be able to provide insights into the variations across program sites (if multiple sites are funded) in terms of environmental factors (e.g., the community), sponsoring organization characteristics, types of fathers served, service delivery strategies, program components, and the relative effectiveness of the differing strategies employed by each site. The discussion guide included in Appendix C can be helpful in structuring interviews with administrators at funding and oversight agencies, though this instrument will need to be tailored to each specific interviewee.

5. Program Participants and Individuals Not Participating in the Program

During the process evaluation, evaluators should conduct semi-structured interviews with randomly selected fathers (and other individuals) who have and have not participated in responsible fatherhood program activities. In contrast to the more structured and larger sample surveys that might be conducted as part of the impact evaluation, these interviews should be less structured and should involve probing of participants and non-participant views on the responsible fatherhood program and its effects. If possible, participants and non-participants should be interviewed individually; if not, they should be interviewed in small focus groups (with 5 to 7 individuals). Interviews with participants could be structured using questions from the discussion guide found in Appendix C.

Participants should be asked about how they first heard about the responsible fatherhood program, why they decided to join and stay with the program, which activities have been most (and least) helpful, and what types of services they felt were missing but needed. They can also provide anecdotal information about their experiences with the program and how it has helped them to overcome problems. They may also be able to describe ways in which their family and other participants were (or were not) assisted by the program.

To supplement the information collected through interviews with participants, it would also be important to conduct interviews with the mothers of participants' children. Such interviews would provide valuable information about how the mother, children, and other family members may have been involved in and affected by services received through the responsible fatherhood initiative. Such interviews would also provide an interesting point of comparison with the perspectives of program participants (e.g., do the views of the father and mother coincide with respect to the effects of the program on the father's relationship with the children).

Another possible source of information would be participant self-evaluations that could be completed at various points in each participant's involvement in the program. For example, such self-evaluation could be completed at the end of receipt of a specific service (e.g., at the end of an eight-week parenting class) or at periodic points if a service is ongoing (e.g., at three-month intervals as an individual proceeds through one-on-one counseling). Participants could be asked to rate the quality of services received (e.g., on a five-point scale), the effects the services had on themselves and their families, and suggest ways in which services might be improved. Such information would be valuable both from the standpoint of evaluating the program and providing rapid feedback for improvement of individual program components. Inclusion of such information in the automated data system would be helpful to both program managers and the evaluator.

Non-participants may be able to provide additional background information on their neighborhood. They should be able to describe some of the types of problems they face at home and in their community. If they have heard of the responsible fatherhood program, they can also explain what they think it is, how it is perceived within the community and among other fathers, and why they are not participating in the program.

6. Community Leaders and Residents

Responsible fatherhood initiatives are expected not only to improve the lives of program participants, but also to affect their families and communities. As a result, it will be important to interview community members. Similar to interviews conducted with participants, interviews with community members should include many open-ended (versus close-ended) questions and probing of respondents. Some of these interviews (maybe one-third) should focus on community leaders (e.g., religious leaders, local politicians, members of local neighborhood associations, etc.). The other interviews should be with randomly selected members of the local community. Appendix C contains a discussion guide that illustrates some of the questions that could be asked of community leaders and/or residents.

In general, community leaders and residents should be able to provide contextual information about community problems and service needs. They may also have knowledge about other programs that exist or have existed in the community and reasons for their successes or failures. Interviews with community leaders and residents can also be useful for obtaining information on the extent of knowledge about the program and its objectives among community residents. Community leaders and residents familiar with the responsible fatherhood program may be able to provide some insights on how the program was implemented within the community and whether the program has had any demonstrable effects on participants, their families, and/or the surrounding community. Even if they are not aware of the program, community residents may be able to suggest ways in which the program can be more responsive to community needs.

B. Secondary Data

There are two major types of secondary data that should be collected as part of the process evaluation: (1) information that exists in program documents; and (2) data collected as part of a client management information system. These two sources of information are discussed below.

1. Existing Documents

Case Files: Responsible fatherhood program sites are likely to have a case file system in place, which includes a series of written forms for assessing and tracking program participants. The sites we visited in developing this evaluation design maintain a number of forms and written notes on each participant in their program. For example, in one program, a short (one page) intake form is completed usually during an initial in-home visit to a potential participant. This form captures some basic demographic data about the individual -- age, ethnicity, marital status, last grade completed, employment status, legal concerns, and several other items -- as well additional data about other family members (e.g., name, whether paternity has been established, relation, date of birth, and address/telephone number). Other forms used by this program focus primarily on establishing participant goals and action steps needed to achieve the goals, and monitoring progress toward the goals. These forms include mostly handwritten notes (and could not be entered into an automated data base, except perhaps in the form of a text file). The number of contacts and hours of counseling is maintained for each participant (on a daily and monthly basis). In addition to the forms described above, case managers and counselors maintain narrative notes within case files that document discussions with fathers and other family members (particularly during counseling and case management sessions) and recommended courses of action.

As part of the process evaluation, the evaluator should review a randomly-selected sample of case files at each site. The narrative notes maintained in case file records are revealing of both the wide variety of problems encountered by participants and the courses of action taken in response to problems by case managers and participants. A case file abstraction form might be used by the evaluator to systematically abstract (and analyze) client problems, recommended solutions, and determine the extent to which clients demonstrated improvement.

Statistical Reporting and Other Program Documents: Data on levels of program participation and service provision may be maintained by each site and submitted in the form of a monthly, quarterly, or annual progress report to funding agencies. Such reports may begin with a written summary of the site's program activities for the reporting period. The narrative portion (if one exists) is likely to provide a history (e.g., month-by-month record) of implementation experience at each site, including issues such as staff turnover, space constraints, and coordination problems.

The report may also provide statistical information on client characteristics and service delivery (e.g., monthly counts of the number of participants receiving counseling services). The progress reports should be collected and reviewed for each site in the evaluation. If they extend back before the evaluation, they can provide background on how the program evolved and changed over time, as well as a baseline of statistical data against which it may be able to analyze current service levels and outcomes.

However, because of likely changes in the reporting formats for statistical data (over time before the start of the evaluation effort) and a lack of consistency in the methods for collecting and reporting statistical data across sites, it is not recommended that the statistical portion of these reports be used as a source of data on program participation or services. In general, before the statistical portion of these progress reports could contribute to the process evaluation (e.g., to show trends in service utilization), a standardized reporting format is needed, along with regular quality control checks to make sure that standard definitions are being used across sites (e.g., what constitutes a participant or receipt of service by a participant). Quality control of report data is essential. If possible, the evaluator should design (during the design phase of the evaluation effort) and implement a standardized monthly progress reporting system (backed up by individual client records) that each site can use throughout the evaluation period.

In addition to progress reports, each program is likely to maintain (in varying degrees) other program documentation, such as their original proposal(s) for funding, directives from funding agencies, pamphlets and flyers, memoranda, and other planning documents. All of these may be helpful to the evaluator in describing the design, start-up, and ongoing operations of the program.

2. Client Forms and Management Information System (MIS)

A potentially valuable data source for the evaluation effort (as well as to support day-to-day program operations and reporting) is a comprehensive and valid automated system of client records. An evaluation of a responsible fatherhood program (as well as day-to-day operations of the program) can be greatly facilitated by the development of a comprehensive participant data system. It should be noted that such a data system could be developed prior to the initiation of a process evaluation and is an important management tool for programs to develop even if a process or impact evaluation is not undertaken. The sections below provide a suggested outline of an automated participant management information system (MIS). The discussion begins with a description of manual forms that might be completed by responsible fatherhood program staff. This is followed by a suggested model for an automated MIS that could be used by each site to track program participants. The system should be designed, to the extent possible, to: (a) minimize implementation costs; (b) minimize the burden of data collection and entry for site staff; (c) provide case managers with client level data for assessing client risks and long-term tracking of client caseload; (d) collect data that will permit objective analysis of client characteristics, risk factors, and outcomes; and (e) track types of services received by each client.

a. MIS Forms

To ensure high quality and complete data are collected on clients and to assist case managers in the delivery of services, a standardized set of client forms should be developed that tracks participants from the time of intake to the responsible fatherhood program to the time of exit and, if possible, beyond, for a year or longer. Examples of the types of forms necessary include the following:

Intake Form: An intake form should capture basic demographic characteristics and other relevant background information to be used by intake workers for eligibility processing and to begin developing a client record on each potential participant. This form should be completed during the client's first or second contact with the program. It should not be so long or burdensome that it is deterrent to participation in the program.

Assessment Form: An assessment form is both helpful for case managers and counselors in formulating strategies for assisting participants and for providing useful information for the evaluation of the responsible fatherhood initiative. This form should be completed when the individual is enrolled in case management services and during their first several contacts with the case manager.

Service Utilization Form: This form is used to track the services received by participants on a monthly or quarterly basis. It should be completed by case managers for each participant within his/her caseload.

Outcome Form: This form is used to document participant outcomes (e.g., establishment of paternity, completion of education or training programs, finding a job, etc.). Data should be entered onto this form periodically (e.g., quarterly) or at a minimum at the time the participant exits from the program.

Although the forms used by each program do not have to be identical across sites, it is strongly recommended that sites maintain at least a core of similar data on participants, their families, risk assessment, service utilization, and outcomes. Without some degree of conformity, it is difficult for evaluators to use MIS data to make relevant and valid comparisons across sites.

b. Suggested Design of a Client MIS

It will be necessary for the evaluator to work closely with the responsible fatherhood site(s) on the design, development, and implementation of a client MIS system. In a multi-site evaluation effort, it is recommended that an advisory committee be formed that would include representatives from each site included in the evaluation effort, the evaluator, and other personnel with expertise in PC-based data systems. This group should work collaboratively on the development of a system that will effectively meet the operational, reporting, and evaluation needs of all parties.

Because staff time and energy is expended on developing and maintaining the MIS (e.g., completing and entering client forms), it is imperative that they get some type of "return" for their efforts. For example, the system should assist case managers with both assessment and better tracking of participants, as well as reduce duplicative entry of data and manual counts for (monthly/quarterly) progress reports. Hence, the MIS should include a report generating capability that enables program staff to easily generate aggregate monthly statistical reports and other reports on clients to suit their needs.

Data Files and Entry Formats: Once there is agreement on a set of forms, it is necessary to design and test data files and data entry formats. There are a variety of different data base software packages that can be used to automate the system (e.g., DBASE, FoxPro). Whatever data base package is selected, it should be sufficiently transferable to other applications, such as software for conducting statistical analyses. If a multi-site system is developed, each site should be able to enter and edit data, sort/index data, delete individual records, and print out reports. The data structure and data entry screens should be set up so that they can be easily altered to customize the application for the sites at the time the system is installed, or to add new data elements or additional forms in the future. In addition, the system should be designed so that each site can create their own supplemental data files, which can be easily matched with the core MIS data file (using a unique client identifier, such as Social Security Number or a client ID Number).

Reporting: The report generating software used will depend on the software selected to operate the system. The report generating software should allow users to both print out aggregate (summary reports) as well as reports showing individual data on clients. This enables sites and the evaluator to monitor the quality of the client data files and to verify aggregate statistical reports submitted by sites summarizing the number and characteristics of fathers served, types of services provided, and outcomes. There are a number of low-cost and highly-flexible report generating programs available for this purpose.

Computer Hardware and Software: Sites may need to upgrade their existing computer hardware or software to operate the MIS. If needed, the evaluator should help with selection of equipment to ensure that it is compatible with the automated MIS application that is developed. In addition, the evaluator could assist sites with the purchase of statistical and/or graphics software that sites could use for their own analysis efforts.

V. Analysis and Reporting

The next two sections illustrate a potential approach to (a) describing and assessing results of a process evaluation of responsible fatherhood initiatives, and (b) how these process evaluation results and implications should be reported. In addition to supporting the overall evaluation effort, the results of the process evaluation could be used as feedback to assist sites in making operational changes to enhance program performance. For example, the process evaluation can be important in helping sites to draw from the experiences of one another and in providing helpful feedback on how to better target services on specific needs of program participants. The discussion below assumes a multi-site evaluation design. Process evaluations of several different programs for a multi-program evlauation alterations or for a single-site would involve many of the same sorts of analyses.

A. Analysis

There are several levels of analysis that should be conducted as part of the process evaluation. Analysis should begin with careful analysis at the site level, move on to comparisons across the responsible fatherhood sites, and conclude with a synthesis of the findings across sites. In a multi-site evaluation, it is important to document whether there are significant differences in the characteristics of the sites that might effect program outcomes. For a single-site evaluation it is also important to understand factors that will affect program outcomes, but it will not be necessary to determine how those factors differ across sites.

Descriptive Analysis of Each Site: The first step in analyzing data involves examining the responsible fatherhood program results at the site-level. Without a thorough understanding of each site's experience, the overall evaluation effort is likely to fail. Therefore, the analysis effort should begin with the development of a case study report on each site. The case studies should be based on client-level data, site visits and interviews, project-level documents and reports, and other sources of information on each included in the evaluation. Each case study should include a complete description of the project design, start-up activities, organization of the program, types of fathers and families served (and not served), types of services provided and the delivery system, and subjective assessments of the benefits and costs of the approach.

Comparative Analysis and Synthesis of Findings Across Sites: Once the site-level analysis is completed, the evaluator should conduct a comparative analysis across sites. The site-level analysis should provide much of the information that is necessary for both generating cross-site comparisons and for synthesizing results across sites. This type of analysis might include cross-site comparisons along the following dimensions:

- characteristics of the sites;
- trends in program participation and characteristics of participants;
- levels of service and assistance provided for program participants;
- participant outcomes;
- program costs; and
- program linkages.

For example, systematic comparisons of the characteristics of each of the sites included in the evaluation will be important. Areas of comparison across sites might include relative funding levels, types of services/activities provided, and outreach and recruitment efforts. For some characteristics (such as funding levels, participation, and date of initiation) it may be possible to make quantitative comparisons. In other areas -- for example, specific services offered to participants or problems encountered in program start-up -- the comparisons will involve more qualitative assessments. The assessments in this area should be rich in narrative comparing and contrasting the design features of the demonstration sites.

It should then be possible to compare the characteristics of program participants across demonstration sites. For example, comparisons can be made across basic demographic characteristics (e.g., age, race/ethnicity, marital status, levels of education achieved, etc.) and selected background factors that might affect participant outcomes in the program (e.g., past patterns of employment, use of illegal drugs and alcohol, criminal record, etc.). The evaluators can generate frequencies from both aggregate data submitted by each of the sites and client-level data collected in the MIS. The advantage of working with the client-level data is that it should be possible to analyze the types of fathers served by each site. For example, it should be possible for the evaluator to cross-tabulate age, race, and a variety of risk factors of participants to describe the types of fathers that have been served by each site.

Next, the analysis effort should assess the types of services provided and received by program participants. For example, comparisons might be made of the percentage of participants within each site that received each type of service. A further step, if data are available, might involve comparisons of the average hours of assistance received within specific types of service categories (e.g., individual and family counseling).

Finally, evaluators should examine gross outcomes for program participants. These analyses will set the stage (and provide some preliminary findings) for the more elaborate and controlled analyses planned under the impact evaluation. Cross-site analysis in this area should begin with a comparison of relative frequencies on a range of key outcome variables. Such an analysis should provide some clues about the impacts of programs, although it will have limited value in explaining whether successes or failures are the result of the types of fathers that are served, environmental factors, or the site-specific intervention. Analyses of participant outcomes could be done on a variety of measures, such as establishment of paternity, fathering new children, quantity/quality of interactions of fathers with children, changes in educational attainment, patterns of employment, incidence of incarceration and criminal activity, and use of alcohol and illegal drugs.

At this point, it may also be possible to begin to examine (as part of the process evaluation but leading to the impact evaluation) potential relationships that may exist between changes in participant outcomes and (a) participant demographic characteristics, (b) measures of participant risks, and (c) involvement in various interventions within each site. For example, it may be possible to conduct a cross-site comparison of paternity establishment by selected participant characteristics.

Some crude analyses of the costs of program services might also be possible based on data collected during the process evaluation. For example, it may be possible to compare the costs of providing specific types of services (e.g., counseling sessions), across responsible fatherhood sites, on a per-participant basis.

Overall, the comparative analysis, which is likely to depend primarily on qualitative assessments, frequencies, cross tabulations, and standard statistics (e.g., mean, median, ranges), should set the stage for more elaborate explanatory analyses that would be conducted as part of the impact evaluation. Results of the comparative analysis across sites must be interpreted with caution.

Responsible fatherhood program sites will be serving participants with varying characteristics in different economic, social, and policy environments. The process evaluation will not provide controls (e.g. comparison group data) sufficient to allow inferences about the impacts of the program on outcomes for participants.

B. Reporting

Two closely related final reports can be developed as a result of the process evaluation: a synthesis report and a case studies report. The first report can serve as an overall (process) evaluation of the responsible fatherhood sites studied, while the second report provides detailed case studies of each program site, which may facilitate the replication of successful aspects of responsible fatherhood initiatives in other locations.

The final synthesis report should provide full documentation on the study, including: an executive summary, objectives of the study, the evaluation methods used, analyses of interview and site visit information, analyses of site-level and participant-level data files, and related findings, conclusions, and recommendations. Several appendices (or working papers) might accompany this report -- for example, providing documentation on data bases developed, background on data collection instruments, and procedures used during the site visits. An important part of the final report should address the policy implications for future development of responsible fatherhood programs. Preliminary options for promoting effective strategies for assisting non-custodial fathers (and their children) should be developed. For each option identified, any restrictions on funding levels, required matching funds, or eligibility should be noted where applicable. In addition, study findings should be used to develop or support the options that are identified.

The second report -- case studies of each of the sites evaluated -- should convey the design, ongoing operations, delivery system, types of fathers served and not served, participant outcomes, and program costs. This report could provide a separate chapter on each responsible fatherhood site evaluated. Each case study should be structured similarly (although each may include different sub-sections) and organized so it can stand on its own (e.g., a case study of a site could be reproduced for dissemination as a stand-alone document).

VI. Conclusion

This chapter outlines a basic design for a process evaluation of responsible fatherhood programs. The design addresses a series of evaluation questions aimed at understanding, among other things, how the program was implemented, what assistance/services it provides, and who it serves. The design also considers changes in participant outcomes, which sets the stage for a more in-depth impact analysis. It is suggested that a process evaluation be started as early in the overall evaluation effort as possible. By starting early, evaluators will be able to begin to provide feedback to the sites so that they can better target and refine their service delivery. Early implementation of the process evaluation will also support the impact evaluation component by providing information that may be used to develop the methods for sampling, data collection, and data analysis.

It warrants repeating that the effort required to implement a full-fledged process evaluation is substantial. Given limited resources, an important first step is narrowing the scope of the evaluation questions. Selecting the questions that are most important to the stakeholders will provide focus to the data collection and analysis activities, and thereby reduce the resources required.

Return to ToC

CHAPTER THREE

MAJOR DESIGN ALTERNATIVES FOR AN IMPACT EVALUATION

I. Introduction

In this chapter we discuss two major design choices that must be made in the planning process for an impact evaluation. These choices concern: whether to use an experimental (i.e., randomized program assignment) or non-experimental design, or some hybrid; and whether to evaluate each individual site independently or to pool the data from multiple sites and evaluate them jointly. Given our limited knowledge of fatherhood programs as well the resources that might be available to evaluate them, it is not appropriate to recommend which alternatives to select for an evaluation. Instead, we describe the options and discuss criteria to be considered in making the choices.

The criteria we discuss include:

- **Feasibility** -- Are there technical, ethical, logistical, or other problems that would make implementation of the design feature inappropriate or problematic?
- **Impact Estimator Bias** -- For each design alternative, what are the potential sources of bias in estimates of program impacts on outcomes, and how substantial is bias likely to be?
- **Estimator Precision** -- How will the design feature affect the likely size of random estimation error?
- **Cost** -- What are the implications of the design feature for the cost of the evaluation?
- **Other** -- Does the design feature add or detract from the quality or value of the evaluation in any other way?

We conclude the chapter with a summary of the most important points with respect to these criteria for each design feature.

II. Experimental vs. Non-Experimental Designs

A rigorous evaluation will require a "treatment" group -- the group that receives program services -- and a control or comparison group of some sort. In this section, we present three alternative designs for these two groups. The three designs are: a classic, experimental design, with randomized assignment to treatment and control groups; a non-experimental design that uses a non-randomly selected group of fathers who do not receive program services as a comparison group; and an intermediate design that we call the "randomized outreach" design. The last design is a modified experimental design that preserves enough of the experimental design's features to address what is likely to be the most problematic aspect of a non-experimental design, the bias in the estimates due to unobserved differences in the treatment and comparison groups, but also avoids some of the problems inherent in the experimental design.

We describe each design below, and discuss its strengths and weaknesses. Which design is best for an evaluation depends on both characteristics of the program and on the resources available for the evaluation. We discuss criteria for selecting among the three designs at the end of the section.

All three designs would use the same primary data collection methodology -- a baseline survey with at least one follow-up -- for both the treatment and control or comparison groups. For the treatment group, the baseline survey would collect information about the characteristics of fathers -- including their relationships with their children -- before program participation, while the first follow-up would collect outcome data and information on study participants' receipt of services from other programs shortly after the father has completed the program. For the control or treatment group, the baseline and follow-up surveys will collect the same information at comparable points in time. Details of the data collection plan appear in Chapter Six.

One other common feature of all three designs deserves mention here. Before selection of treatment and control or comparison group subjects, the evaluators would identify volunteers from specified populations to participate in a "long-term study of non-custodial fathers," not telling them that the purpose was to evaluate a particular program, and would offer incentives for volunteering to participate in the baseline and follow-up surveys. The purposes of this feature are: to reduce differences in the measured outcomes for the treatment and comparison or control group that are due to differences in the willingness of fathers to volunteer for, and complete, the study; and to disguise the fact that success of a particular program will be judged, in part, on the basis of their behavior.

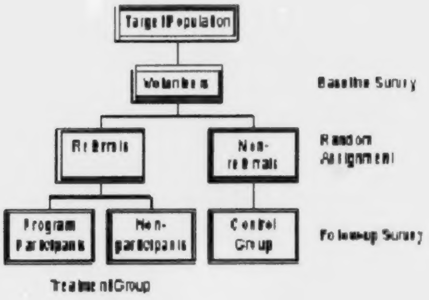
A. Experimental Design

Target Population

The experimental design (***Exhibit 3.1***) begins with the identification of the target population for the evaluation -- the population of fathers that the program targets for service. In general, this is the population of non-custodial fathers in the community that is served by the program, but it may be defined as the population of non-custodial fathers who come in contact with one or more recruitment or referral sources. An example is the maternity ward at a local hospital, in which case the target population is, at least in part, the non-custodial fathers of infants born to unwed mothers in that hospital who are contacted by the referral sources. Other sources may include: fathers of children participating in a local welfare program, fathers residing in a specific geographic area, or fathers who are incarcerated.

Exhibit 3.1

Experimental Design



Study Volunteers

To conduct an evaluation, it is necessary to contact fathers in the target population and ask them to volunteer to participate in a study of non-custodial fathers and their children. This could be most easily accomplished by an outside referral source that, in the absence of the evaluation, would be in contact with the same fathers and would refer them to the program. The referral source would be asked, instead, to refer the father to the researchers conducting the study.

It may be necessary to offer fathers an inducement to participate in the study in order to obtain an adequate number and mix of volunteers. This could be a payment for responding to the baseline survey. An alternative is to give them an item that would benefit the child, but this might result in fewer volunteers per dollar spent on incentives and would skew the mix of volunteers towards those that are most motivated to benefit their children.

Baseline Survey and Random Assignment

Fathers who volunteer would then be contacted by the evaluators for the administration of the baseline survey. Following completion of the survey, the evaluators would refer randomly selected fathers to the program. These fathers would constitute the treatment group, and those not referred would be the control group. All fathers completing the baseline survey would be asked to provide the name, address and telephone number of at least one contact person -- individuals who "always know how to contact the father" -- so that they may be included in the follow-up survey. An incentive for completing the follow-up survey may be necessary to obtain a high participation rate, and the father should be informed of that incentive at this point.

Treatment group fathers would not necessarily participate in the program just because the evaluators refer them to the program. The evaluators could ensure a high participation rate among treatment group fathers by several means. First, a screen could be used to screen out potential study volunteers who are very unlike the program's participants. Characteristics of fathers that are rarely or never observed among participants could be determined with the assistance of the program, and used as the basis for the screen. Questions concerning the fathers interest in obtaining specific types of assistance might also be asked. Fathers identified as unlikely to participate would be screened out of both the treatment and control groups. Screened out fathers could be dropped from the study entirely, or the data collected from them might be used for an auxiliary, descriptive analysis.

Second, with the permission of the father, the evaluators would help the father get in touch with program staff, who would then use any means at their disposal to encourage participation. Note that any methods used to encourage participation of referred fathers become part of the treatment, because they are not offered to control group fathers.

An Alternative Experimental Design

An alternative experimental design that might achieve higher participation rates among treatment group members would ask fathers to volunteer for program participation *before* random assignment. This would screen out all fathers who, at least initially, did not want to participate. It would not, however, guarantee participation from all treatment group fathers because some might change their mind at a later date. Further, it would not give the program an opportunity to encourage participation by fathers who might otherwise not participate. This assumes that participation is largely voluntary. In a situation where fathers are "required" to participate, perhaps by court order as a condition of parole or visitation, this would not be an issue.

Another problem with the alternative approach is that control group fathers would be made aware of the program, would likely be disappointed at their assignment to the control group, and might significantly change their behavior as a result of that assignment. Further, both treatment and control group fathers would know that they are part of a study to evaluate the program, which might also change their behavior, whereas under the recommended approach volunteers would only be told that they are participating in a study of non-custodial fatherhood. These last two problems can also arise under the recommended approach, but to a substantially lesser degree.

Follow-Up Data Collection

Follow-up data should be collected from as many study volunteers as it is feasible to reinterview. Follow-up data should be collected after a specified interval following the baseline interview. The length of the interval should be long enough so that those who participate in the program are likely to have completed their participation, but not so much later that participants are likely to have forgotten significant information about their participation in the program or about immediate post-program outcomes. It is necessary to define a fixed interval after random assignment, rather than interview participants shortly after they complete the program, so that data collection for the control group will be comparable to data collection for the treatment group.

It is important to collect follow-up data from such "non-participating treatment" fathers -- fathers who were referred to the program but did not participate -- in order to adjust for estimation bias that would likely result if participating fathers alone were compared to control group fathers. Participating fathers are a self-selected subset of all treatment group fathers, and are likely to have a higher probability of positive outcomes than the average control group father even in the absence of participation.

Data Analysis

A full discussion of data analysis is deferred until later in the report (Chapters Seven and Eight). The discussion here is intended to indicate the nature of the analysis and to provide background for the discussion of criteria for selecting a design that appears later in this chapter.

Analysis of the data under an experimental design can be very simple because random assignment would eliminate all but chance differences between the baseline characteristics of the treatment and control groups. Differences between treatment and control group means of the outcome variables from the follow-up survey are the simplest measures of the program's impact.(1) There are, however, several reasons to use more complex analyses.

First, we assume that some treatment group members would not participate in the program. Difference in means estimates that exclude non-participating treatment group members from treatment group means likely overstate the effect of participation, due to the self-selection problem mentioned above. If, instead, non-participant treatment group members are included in calculating the treatment group mean, the difference in means is likely to understate the impact of the program for those who actually participated.

A simple way to obtain an unbiased impact for those who participated would be to divide the difference in mean outcomes between all treatment group and control group fathers by the proportion of treatment group fathers who participate. This approach follows from the expectation that the sample means for the treatment group will satisfy the following equation:

treatment mean - control mean = participation impact x % participating

where "treatment mean" refers to the mean of an outcome variable for the full treatment group, "control mean" is the corresponding mean for the control group, the "participation impact" is the percent effect of participation on the outcome variable, and "% participating" is the share of treatment group members who participate.(2)

Second, some members of both the treatment and control groups will drop out of the study, and there may be substantially more attrition from the control group than from the treatment group because of the latter's participation in the program. Hence, it may be important to study the determinants of attrition, and to make adjustments for differences in attrition between the control and treatment groups.

Third, more precise impact estimates can be obtained by controlling for characteristics of treatment and control group members that are measured in the baseline survey, including baseline outcome measures. Multivariate analyses that incorporate this information can be explain a significant share of the random variation in outcome across fathers, reducing the size of any remaining difference between treatment and control group means that is plausibly due to chance. This analysis would also include a multivariate analysis of program participation among treatment group members. As we discuss further in Chapter Seven, the results of the participation analysis would be interesting in their own right, as well as useful in improving the quality of the impact estimates.(3)

Example of a Possible Experimental Design

One of the programs we visited, the Racine Goodwill Industries Program, primarily serves fathers who are "referred" by the court system for failure to comply with child support orders. Services include employment services that are provided through an arrangement with another organization as part of Wisconsin's Children First Program, and a variety of other services, such as parenting and fatherhood responsibility courses, that are provided by Goodwill (see Appendix B for a more detailed description). Approximately 50 to 60 fathers are referred by the courts each month.

A randomized evaluation of the employment service component of the program alone is currently being conducted by the State as part of an evaluation of the Children First Program. This represents the only effort of which we are aware to formally evaluate a specific component of a fatherhood program. It serves as an example of an experimental design approach and illustrates some of the kinds of issues that fatherhood programs will face in conducting impact evaluations.

For this evaluation, fathers who are sent to the Goodwill by the court are randomly assigned into control and treatment groups. Treatment fathers receive employment services as well as other services provided by the program, while control fathers receive the other services alone. Thus, this evaluation focuses on the impact of the employment services *conditional on receipt of the other services.*

Preliminary findings from the State's evaluation were provided to us by the State (see Appendix D). They show that the mean child support paid by treatment group fathers increased by 76 percent from the six months before referral to six months after referral, while mean payments from control group fathers increased by 62 percent. By the second six months after referral, mean payments from control group fathers outpaced those from treatment group fathers -- up 82 percent from the six months before referral compared to 77 percent for treatment fathers. Other related outcome measures (number of payments made and number of fathers making payments) show similar findings.

There are several possible explanations for these results. One is that court enforcement *per se,* rather than services provided, account for improved payments. Another is that the fatherhood services provided by Goodwill, rather than the employment services, are the critical determinant of increased support. The latter conclusion is discounted by the fact that pre-post increases in support payments for the Children First Program in other Wisconsin counties appear to be as large as in Racine during this period, but these counties do not provide services that are comparable to fatherhood services provided by the Goodwill Industries Program in Racine.

Another explanation of the small differences in results for the treatment and control groups is possible spillover problems. The Goodwill counselors knew who the control subjects were, and, as reported to us, were uncomfortable with denying the subjects with services that they thought would be beneficial. The counselors faced an ethical problem, and the immediate needs of their clients may understandably have taken precedence over the evaluation's needs. While the counselors could not send their clients to obtain the employment services, they could provide compensating services.

It might be feasible to conduct the "reverse evaluation" by random assignment -- an evaluation of fatherhood services conditional on receipt of the employment services -- although there may be institutional obstacles to such an evaluation. This evaluation would show whether the package of fatherhood services provided directly by Goodwill "adds value" to the employment services. The reverse evaluation would also examine a broader range of outcome measures, rather than focusing on child support. A spillover problem could arise here too. To reduce this problem, the courts might refer the control subjects -- those receiving employment services only -- directly to the provider of those services, avoiding contact with Goodwill Industries staff. Of course, staff providing employment services might find themselves in the same bind as Goodwill staff did in the evaluation of the employment services.

An experimental evaluation of the combined services might be more useful to program funders, but would be more problematic. According to the child support enforcement office, the alternative to assigning fathers to the program is sending them to jail, something they are prepared to recommend! Further, the program can accommodate all fathers who are currently referred, so the program's manager is not willing to deny services to fathers who would otherwise be clients.

Discussion

It may not be feasible to implement an experimental design. A randomized design would likely require cooperation from referral sources -- including asking them to *not* refer some clients who might otherwise be referred. Referral sources and others are likely to object to this on ethical grounds because some fathers would be denied services that, in the absence of the evaluation, they might receive. This is especially likely to be true if the program has the capacity to accommodate all referrals.

While many sources of estimator bias that are avoided with an experimental design, potential bias remains because the study is not "blind." Program staff are likely to know their program is being evaluated, and it may behave differently as a result. Study volunteers, staff at the referral sources, and others may also learn about the purpose of the study and also alter their behavior to influence the outcome.

The "non-blind" nature of the study will be a problem for any of the designs we are considering, but may be more problematic for this design than for a non-experimental design because treatment and control subjects may be likely to come in contact with one another -- they come from the same target population and are in contact with the same referral sources. "Spillovers" -- information obtained by control group fathers from treatment fathers, competition between control and treatment fathers, disparagement of the treatment by control group fathers, alternative services obtained by control group fathers, etc. -- will all affect impact estimates.

The size of the program to be evaluated may be too small to generate a sample size that is large enough to yield sufficiently precise estimates. Based on the two programs we have examined to date, the evaluators would be fortunate to obtain 200 subjects from a single site over a one-year period. If 100 were assigned to treatment and 100 to control, a simple difference in percent would have to be at least 12 percentage points to be statistically significant at the five percent level.(4) This can be improved upon to some extent by using multivariate methods, but the estimates are likely to be inadequately precise for many purposes with groups of this size. Other options to improve precision would be to pool data from multiple sites or extend the sample collection period, both of which may have other problems. Problems with pooling data from multiple sites are discussed later in the chapter. Lengthening the sample collection period would delay completion of the study and would increase the chance that the evaluation would be compromised by significant changes in the program, its environment, or the evaluator's staff.

Any high quality impact evaluation will be costly. For an experimental design, significant cost sources will include: developing a detailed plan, including instruments; implementing the methods for soliciting volunteers; conducting the baseline survey and randomly assigning them to treatment and control groups; maintaining contact with study participants and conducting the follow-up survey; preparing the data; analyzing the data; and disseminating the findings. Except for costs incurred to randomly assign volunteers, the costs for each component would likely be no larger than they would be under alternative designs.

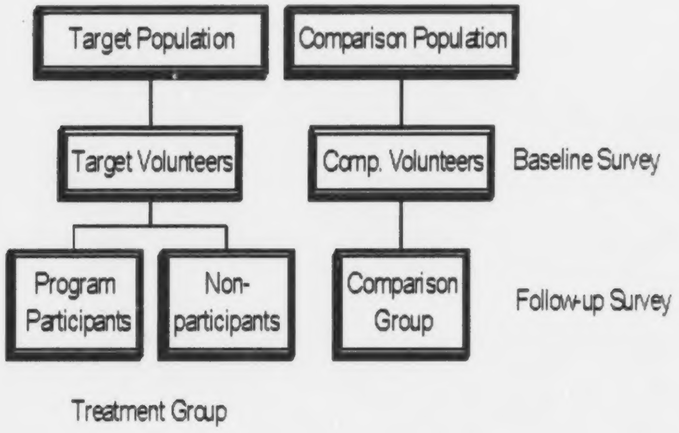
B. Non-Experimental Design

Target Populations

For the non-experimental design, the evaluators would identify two distinct separate treatment and comparison group target populations (***Exhibit 3.2***). The treatment group target population would be for the population served by the program to be evaluated -- the same population that would be the target population for the whole evaluation under an experimental design. The comparison group would be a population that is not served by the program or a comparable program, but is otherwise very similar to the program's target population. Thus, for instance, if the target population for the program is non-custodial fathers of newborns at a specific hospital, the target population for the comparison group could be non-custodial fathers of newborns at one or more similar hospitals that are not served by the program or a comparable program. If instead, the target population is non-custodial fathers within a specific geographic area, the comparison population would be the corresponding population in a geographic area that is similar in socioeconomic characteristics.(5)

Exhibit 3.2

Non-Experimental Design



Study Volunteers

Under this design it would be necessary to solicit study volunteers from both treatment and comparison populations *in an identical way*. The purpose of identical solicitation is to obtain two sets of volunteers that are as comparable as is feasible. Informing potential volunteers from the treatment population that they will have an opportunity to participate in the program is likely to get a set of volunteers that differs from the comparison volunteers in a way that is difficult to measure -- related to the volunteers' desire to participate in the program. Incentives to participate in the study might be required to obtain a desirable number and mix of study participants, just as in the experimental design.

Baseline Survey

As in the experimental design, a baseline survey would be conducted by the evaluator once contact is established with the study volunteer. Following the completion of each interview, the respondent would be asked to keep in touch through a contact person, and to eventually participate in a follow-up survey. Respondents from the treatment population would all be referred to the program, through the same process used for randomly assigned treatment fathers in the experimental design.

Follow-up Data Collection

Follow-up data would be collected in the same manner as was described under the experimental design, including data for volunteers from the treatment population who elect not to participate in the program.

Data Analysis

While differences in means of outcome variables could be used to estimate program impacts, such estimates are likely to be biased because of systematic differences between the underlying treatment and comparison populations. Many such differences are likely to be reflected in baseline characteristics of the treatment and comparison group volunteers. Just as in the experimental design, these characteristics can be incorporated in a multivariate analysis to control for observed baseline differences between the groups.

Collection of high quality baseline data and multivariate analysis of outcomes is more critical for the non-experimental design than for the experimental design because baseline differences between the non-experimental treatment and comparison groups are not just due to chance, may be substantial, and may have a strong association with key outcomes. Even after controlling for observed differences in baseline characteristics, remaining differences between outcomes for the two groups may reflect unobserved differences in baseline characteristics. The main weakness of the non-experimental design is that it is not possible to adjust for those differences which are not observed in the baseline data.

Example of a Non-Experimental Design

It may be feasible to conduct a non-experimental impact evaluation of the Racine Goodwill Industries Program, using one or more other counties in Wisconsin as comparison counties. Recall that the program primarily provides services to fathers who are referred by the courts as a means to increase child support payments. As mention previously, other counties in Wisconsin offer more limited services -- employment services only, under the Children First Program.

The State has already collected data that could be used for a limited version of such an evaluation -- pre-referral and post-referral child support data for non-custodial fathers who have been referred by the courts to the Children First Program.(6) This program is operational in other Wisconsin counties and provides limited employment services to fathers who are referred by the county courts.

An evaluation that would be more in line with the non-experimental design presented here and that would broaden the outcome variables beyond measures of child support would require interviews of fathers who are involved in court actions concerning child support at the time these actions are beginning (i.e., the baseline survey), with follow-up interviews several months later (i.e. the follow-up survey). The baseline interviews are especially important because the characteristics of fathers who are subject to court actions, and the nature of those actions, may differ markedly across counties.

This type of an evaluation would compare the effectiveness of Racine's program to the effectiveness of the Children First Programs that are in place in the comparison county(ies). Hence, the evaluation would be limited to analyzing the added impact of the services provided by Goodwill Industries that augment the "customary" Children First employment services. Note that this is also the limited goal of the experimental design for the Racine program that was outlined in the previous section.

A non-experimental design might also be considered for the Baltimore City Healthy Start Men's Services Program. This program is established in two Baltimore areas, East and West Baltimore, and together they serve from 50 to 100 men each year. The Baltimore site is one of 15 Healthy Start programs nationwide. The fathers who participate in the Baltimore Men's Services are non-custodial fathers who are recruited through their children's mothers; the latter are participants in the Healthy Start program. The program is well established. It is obviously too small to apply any experimental design. The number of fathers served annually is small for a non-experimental design also, and efforts to increase the number served during the evaluation period would be desirable. Alternatively, Healthy Start programs that provide similar services to fathers in other cities may exist and could be evaluated jointly with the Baltimore program if the programs are sufficiently similar.

The main goal of the overall Healthy Start program is to reduce adverse birth outcomes, through increased use of appropriate prenatal, post-partum, and pediatric care. A non-experimental evaluation of the main program is already being conducted, using an adjacent Baltimore area as the comparison site. The comparison area has changed considerably since program implementation, however, and may no longer be suitable as a comparison area, but others may be available.

To find fathers for the study from the comparison area, it would be desirable to recruit them in a manner similar to the manner used by Healthy Start. This will be difficult because there is no set of Healthy Start mothers in the comparison area. One approach would be to use Health Start's methods for identifying mothers, then use the mothers to find the fathers. This is cumbersome, however.

Another potential problem with this approach to evaluating Healthy Start Men's Services is that it would really evaluate the impacts of all Healthy Start services, including the Men's Services, because children and mothers in the comparison areas would not be receiving other Healthy Start services. If, instead, comparison mothers were selected from Healthy Start programs in other cities that do not men's services, the impact of Men's Services alone could be evaluated. Determining whether this is possible would require review of the programs in other cities. Differences in the economic, cultural, and policy climate in Baltimore and other cities would also make this design problematic.

Discussion

It is usually more feasible to implement a non-experimental design than an experimental design. This type of design does not normally have an impact on services that responsible fathers would be getting; i.e., those in the treatment group would participate in the program just as they would or would not in the absence of the evaluation, and those in the control group would presumably receive the same services, if any, that they would have received in the absence of the evaluation.

There may be other challenges to feasibility, however. First, a reasonable comparison group must be found, and it may be difficult to find one that is sufficiently similar to the treatment group before treatment in all important respects. Second, collection of data from the comparison group is likely to require cooperation from agencies that serve the comparison population -- agencies that would refer fathers to the program were the program located in their community. Their cooperation seems less likely than the cooperation of agencies that actually make referrals to the program. Generally, collecting comparable data from members of two different target populations is likely to be more problematic than collecting data from members of a single population, as would be required under an experimental design.

The non-experimental design would be much less vulnerable to the spillover effects that might bias estimates under an experimental design, but bias may be a significant problem for other reasons. The most serious is likely to be differences between the separate target populations from which the two groups are drawn. While baseline data can be used to control for the effects of observed differences in treatment and control group members on outcomes, this will be imperfect. Another source of bias is environmental factors -- the local labor market, other community services, etc. -- which may differ substantially across the two groups. Differences in outcomes may reflect differences in environmental factors. Differences that remain constant throughout the evaluation period can be controlled for by comparing *changes* in outcome variables (i.e., follow-up outcome values minus values) for the two groups, rather than the levels of follow-up values. Changes in environmental factors that are different for the two groups (e.g., labor market improvement in one area, but not the other, or changes in the policy environment) would be difficult to control for in the analysis.

For a given sample size, the estimates from a non-experimental design will be less precise than those from an experimental design, depending on how well matched the two groups are.(7) It is likely, however, that a larger sample size can be achieved over a given period of time because the constraint imposed by the program's size applies only to the treatment group, rather than to the combined treatment and control groups. If the comparison group is the same size as the treatment group, then the sample size is potentially twice as large as for an experimental design with equal size treatment and control groups. Thus, in our hypothetical program that has 200 participants per year, the size of the study sample over a one-year period would be 400, rather than 200. This reduces the size of a difference in percent that is statistically significant from 12 percentage points to eight.

For a sample of given size, it may cost more to collect data under this design than under the experimental design because the volunteers would be obtained from a greater number of sources (e.g., referral agencies or geographic areas). Data collection costs will be increased further if the larger sample size that can be achieved under the non-experimental design is sought. Also, because the importance of controlling for baseline characteristics is more important to prevent bias under the non-experimental design than under the experimental design, the evaluator may wish to put more effort into designing and conducting the baseline survey.

C. Randomized Outreach Design

Random Outreach vs. Random Referral

The randomized outreach design (***Exhibit 3.3***) modifies the experimental design in the following simple way. Under the experimental design, randomly selected volunteers are referred to the program, while those not selected are not referred at all. Under the randomized outreach design, all volunteers are referred to the program, but extraordinary efforts are made to encourage participation of a randomly selected subgroup -- the "outreach treatment" group. For instance, while all volunteers would be offered an incentive to continue to participate in the study through follow-up, those selected for the outreach treatment group might be offered a larger incentive if they also participated in the program. Alternatively, researchers or program staff might: more actively "sell" the program to randomly selected volunteers; contact volunteers a few days after the interview to check if they have enrolled and, if not, encourage them further; offer transportation to the program office, etc.

Exhibit 3.3

Randomized Outreach Design

There are several reasons for modifying the experimental design in this way. First, it addresses the ethical problem that may thwart implementation of the experimental design by giving every volunteer an opportunity to participate. Relative to the existing program, it will not deny or discourage anyone from participating; instead it will provide added encouragement to a subset of potential participants.

Second, the spillover effects that might occur under the experimental design are largely avoided. Control group fathers who decide they want to participate will be allowed to participate, so the potential for rivalry between the two groups is greatly reduced. In fact, members of both groups may be unaware that they have been assigned to one group or the other, or even that the purpose of the study is to evaluation the program. "Blindness" of subject is most likely to be achieved if the treatment is limited to extraordinary follow-up marketing activities that would be difficult for volunteers to detect. Use of special incentive payments would be easier for volunteers to detect and, if detected, have an impact on their behavior.

This modification achieves these advantages over the experimental design but preserves the most important feature of the experimental design: differences in outcomes between the control and treatment groups that are not caused by the treatment are due to chance and will be small if the sample is sufficiently large. Here, however, the treatment is not the program, but rather the outreach.

An additional analytical step is necessary to convert the outcome differences into estimates of program effects, as discussed further below.

Participants and Non-participants

Under this design, a substantial number of control group members will participate in the program. If the randomized outreach is effective in increasing participation, the share of control group members who participate will be smaller than the share of treatment group members who participate. Data must be collected for both participants and non-participants from both groups.

Data Analysis

The outcome analysis under this design would compare outcomes from the participant and non-participant groups, using the randomized outreach feature to correct for bias due to self-selection of volunteers into the participant and non-participant groups. If we assume that the effect of participation on an outcome is the same for all volunteers who participate, we would need to divide the difference between treatment and control outcomes by the difference in participation rates to obtain the estimated participation impact. This follows from the expectation that:

treatment mean - control mean = participation impact x % treatment participation - participation impact x % control participation

where "% treatment participation" is the percent of the treatment group that chooses to participate, "% control participation" is the analogous control group variable, and other variables are as defined previously.(8) The formula presented previously for estimating participation impacts under the experimental design is the special case of this formula when "% control participation" is zero.

As in the experimental model, more accurate estimates of participation effects can be gained through multivariate analysis of outcomes, incorporating control variables from the baseline survey. The outcome analysis would be preceded by a participation analysis that would examine the effect of the randomized outreach method and other variables on participation. The results of this preliminary analysis would be incorporated in the estimation of multivariate outcome models to adjust for self-selection into the program, with a variable identifying which subjects received the random outreach.(9) As in the experimental design, the participation analysis itself would be of interest -- perhaps more so because one objective of the evaluation could be to test the outreach methodology. Further, if the number of volunteers is sufficiently large, two or more outreach methodologies could be tried.

There are at least two threats to the success of this approach that reduce its potential usefulness. First, if the outreach is ineffective, participation rates and outcomes for the two groups will be very similar and the measured effect of the program will be insignificant -- even if the true impact of the program is substantial. Hence, success of this approach requires a treatment outreach that is very effective in comparison to the control outreach.

Second, the program probably does not have the same impact on all fathers, and it may be that the impacts on participants from the treatment group who would not have participated had they received the control outreach are substantially greater or less than those on other participants. On the one hand, these "marginal" participants might be fathers who are motivated by the outreach and not by a strong desire to become responsible fathers, in which case impacts may be small. On the other hand, in comparison to other participants, marginal participants may be fathers who would be least likely to achieve desirable outcomes on their own, in which case impacts may be large. To minimize any potential bias, the evaluators will need to examine differences in baseline characteristics between treatment participants and control participants and investigate whether program impacts are related to these observed differences. Evaluators will not be able to adjust for differences between marginal participants and other participants that are not observed.

Examples of Random Outreach Designs

It might be feasible to conduct an evaluation of the Racine Goodwill Industries program using a random outreach design. According to staff we interviewed, it would not be difficult to find many more fathers to participate in the program in a short period, and the program would welcome an opportunity to reach out to more fathers, even if only some fathers reached are referred to the program.

For this evaluation, the evaluator and the program would cooperate to recruit study volunteers. Recruitment could be accomplished through the many AFDC mothers who are in contact with Goodwill Industries because Goodwill administers the JOBS program in Racine. Alternatively, fathers who are program clients might be employed to recruit other fathers they can contact through informal connections. This might be especially useful for obtaining volunteers from among fathers who are the most difficult to reach.

Volunteers would be asked to participate in the baseline survey. Randomly selected volunteers would be encouraged to participate in the program by the interviewer. Follow-up outreach to these same "treatment group" volunteers could be conducted by the program or program clients. Inevitably some of the volunteers who do not receive the outreach (the control group) will participate in the program, but this is not a threat to the evaluation as long as the outreach efforts applied to the randomly selected volunteers produce a substantially higher participation rate among volunteers assigned to the treatment group.

It should be recognized that this evaluation would not result in estimates of the impact of the program on outcomes for fathers recruited through the program's main referral mechanism -- the courts. Instead, it would estimate the impact of the program on fathers recruited through whatever mechanism is adopted. This estimate may be no less interesting than an estimate for fathers referred by the courts would be, but it must be recognized that results are dependent on the recruiting process.

One interesting "side-effect" of a new recruitment effort might be a reduction in referrals from the courts. This can easily be tested by the evaluator, by comparing the number of court referrals for fathers who are in the control group to the number for those in the treatment group.

A random outreach design might also work for the Baltimore Healthy Start Men's Services Program. This program recruits fathers through mothers who are Healthy Start participants themselves. It may be feasible to recruit a much larger set of fathers by these means to volunteer for a study on non-custodial fathers. Financial or other inducements might be used to recruit randomly selected volunteers for participation in the program. It would be necessary to increase the number of participants recruited well above current levels to make such an evaluation viable, but this may be possible. One advantage this design would have over the non-experimental design for the Baltimore program outlined in the previous section, using fathers from adjacent areas in Baltimore for the comparison group, is that it would evaluate the impact of the Men's Services conditional on the other Healthy Start services, rather than the impact of all Healthy Start Services provided by the Baltimore program.

Discussion

While this design has some very positive features, other factors may make it less attractive relative to other designs. First, as with the experimental design, this design will require some cooperation from normal referral sources, whose help may be needed to implement the randomized outreach. Second, for a sample of given size estimator precision may be substantially lower under this design than under either the experimental or non-experimental designs. How much lower will to depend on the effectiveness of the treatment outreach relative to that of the control outreach -- the less effective, the lower the precision. Severe sample size constraints due to program size or costs would make this design unattractive.

Two factors other than sample size increase the cost of this design relative to the experimental design: the cost of the randomized outreach and some additional complexity in the analysis of the data.

A potentially important advantage of this design over the alternatives is that it provides the opportunity to study the impact of the treatment outreach relative to the control outreach. The evaluator could determine the impact of the treatment on participation and could also determine whether the eventual effect on outcomes. Outreach may be a cost-effective method of improving outcomes for non-custodial fathers. The treatment outreach need not be limited to a single outreach method; multiple outreach methods could be randomly assigned.

III. Single Site vs. Multiple Site Evaluations

Funding may permit evaluation of multiple responsible fatherhood programs in the future, and this design is intended as a road map for conducting evaluations of many different sites. The evaluation of each site could be conducted independently. This would allow the evaluation design and data collection methodology to be tailored to each site's circumstances. Tailoring the evaluation in this way might maximize information gained about each site, but would also make it difficult to compare findings across sites. Under a non-experimental design, the evaluator may want to use a design that does not require a separate comparison group for each site; in the extreme, a single comparison group may be used for all sites.

If the programs are sufficiently homogeneous, there would be a significant advantage to evaluating multiple sites jointly; pooling the data across sites would increase sample sizes and contribute to more precise estimates of program and other effects. This may be especially valuable because the programs we are familiar with are all small, and sample sizes from individual sites will be small unless the evaluation is conducted over a very long period. Sample size constraints are of greatest concern if either the experimental or randomized outreach design are used. Evaluation of homogeneous programs in multiple sites can also provide information about the impact of local environments on the efficacy of the program.

The programs are clearly not homogeneous, however, so it is less obvious that joint evaluation of multiple sites would be advantageous.(10) Heterogeneity across programs has many dimensions. A multi-site evaluation can be designed to accommodate some dimensions of heterogeneity successfully, but not others. We discuss three major dimensions of heterogeneity below: program services, program objectives, and target populations. Of these, heterogeneity in target populations poses the greatest challenge to a multi-site evaluation. We would not rule out joint evaluations of programs with heterogeneous populations, but would urge that caution be exercised before proceeding.

A. Program Services

Differences across sites in the types of services offered by programs can easily be accommodated in an evaluation of sites that are homogeneous in other key respects. The evaluator can easily allow for different programmatic impacts across sites. The evaluator can determine whether differences in impacts across sites are statistically significant, but in general will not be able to determine whether differences are due to specific program features or to environmental factors.(11) If this is the only substantial difference between multiple sites, no matter how many, it would make statistical sense to pool their evaluations because the evaluator can take advantage of the fact that effects of other factors (i.e., control variables) on outcome variables are likely to be similar across sites to improve the precision of the estimates (see Chapter Eight). This may be true for selected groups of responsible fatherhood programs.

If there are a very large number of sites that differ only in services provided, and if their programs can be classified in a meaningful way, the evaluator might also be able to demonstrate that some program features, and/or some local factors, are important determinants of success. This scenario appears unlikely for responsible fatherhood programs, however, because they are small in number and heterogeneous in other respects that are less amenable to multi-site evaluations.

B. Program Objectives for Clients Served

We have observed substantial variation in program objectives for clients across the programs with which we are familiar. This variation is likely to be reflected in the impact of the program on various outcome variables that might be used in an evaluation. For instance, a program that places primary emphasis on helping the father obtain employment is likely to have a different impact on employment outcomes than one that focuses more directly on establishing or improving the relationship between the father, his child, and the child's mother. In an extreme case, an outcome variable that seems an appropriate one for one program, given the program's objectives, may seem inappropriate for another program that has different objectives.

Differences in program objectives alone, however, should not stand in the way of multi-site evaluations. It must be recognized that differing objectives result in variation in services (see above) and are likely to be reflected in variation in measured program impacts across the multiple outcome variables. Effects of other factors (i.e., control variables) on outcomes are likely to be similar across sites, so it makes statistical sense to pool the data, but allow for cross-site variation in impacts.

Although multi-site evaluations of programs with differing objectives may be statistically advantageous, there are some negative aspects of such evaluations. First, the program staff may not want to have the impacts of their programs compared to those for other programs on outcomes they may regard as tangential to their primary objectives. Second, a multi-site evaluation will require collection of common data at all sites, some of which might not be collected from all sites if individual evaluations were conducted. Further, data that may have unique importance to one site might collected for an evaluation of that site alone, but might not be collected for a multi-site evaluation.

C. Target Populations

There is substantial variation in target populations across the sites that we have observed, and this variation is the greatest challenge to multi-site evaluations. Evaluations of programs that have similar target populations may be pooled successfully, whereas pooling evaluations of programs with dissimilar target populations would not be very useful and could be misleading. As in the previous section, target populations are sometimes implicitly defined by methods used by programs to identify and recruit fathers, so similarity in these methods across programs may be required to make joint evaluation attractive.

It might be reasonable, for instance, to pool the evaluations of programs that target non-custodial fathers of newborns, especially if those fathers are from communities that have similar demographic and socioeconomic characteristics, and are identified and recruited in a similar fashion (e.g., through the maternity ward at a community hospital). As another example, it might also be reasonable to pool the evaluations of programs that target all low-income non-custodial fathers in a defined geographic area, especially if the areas have similar demographic and socioeconomic characteristics and if fathers are identified and recruited in a similar fashion. For instance, joint evaluation of the multiple IRFFR sites may be reasonable, although further review of the target populations and methods used to identify and recruit fathers at IRFFR sites may be advisable before making such a determination.

The reason that pooling data from sites with similar target populations is attractive, regardless of variation in program objectives and/or services, is that the effects of other, non-programmatic variables on outcome variables is likely to be similar across populations and pooling the data will improve the evaluator's ability to control for such factors. If, however, the target populations are very dissimilar, the effects of non-programmatic variables on outcome variables may vary substantially across the target populations. There would then be no advantage to pooling, and potential harm. We would be skeptical, for instance, about joint evaluation of a program that targets non-custodial fathers of children in Head Start programs with a program that targets non-custodial fathers who have been identified through the criminal justice system. This would not be as big of a concern, however, if the primary purpose of the evaluation were to determine if the program treatment works equally well in different populations. Information on the characteristics of both populations would still need to collected in order to explain why the impact of the program differed between the groups, if significant differences were observed.

Potential harm from joint evaluations of programs with disparate target populations may occur in several ways. First, the evaluators may unnecessarily constrain their data collection activities in each site because they plan to use the data for a common evaluation. Second, data may not be comparable across sites because of differences in data collection methodologies that must be implemented to accommodate differences in the target populations. Third, the evaluators may pool the data without testing whether the effects of control variables in the multiple populations are similar, which could lead to biased estimates of program impacts. The latter problem can be avoided through appropriate testing, but if sample sizes are small the power of the tests -- their ability to detect important differences in the effects of the control variables -- may be low.

IV. Summary

In the introduction we presented five broad criteria for selecting the major design features. The most important aspects of each design feature with respect to each of the criteria are summarized in ***Exhibit 3.4***.

We recommend that an experimental design be carefully considered before considering alternatives. A carefully implemented experimental design will provide the highest quality findings, and findings that are least able to be challenged. It may be that ethical or practical considerations will make an experimental design unattractive, or that potential sample sizes are too small. The randomized outreach design addresses some of the ethical and practical issues that may make the experimental design unfeasible, while preserving the use of randomization to control for differences in unobserved factors. It, too, may not be feasible or may be too costly. Impediments to implementing a non-experimental design are easier to overcome, and sample sizes obtainable may be larger, but questions concerning the adequacy of controls for differences between treatment and comparison groups are likely to arise.

We also recommend that joint evaluations of multiple sites be carefully considered. While there may be important reasons not to pursue this option, the gains from increasing sample sizes and improving comparability of findings across sites could be very large.

Exhibit 3.4

Summary of Strengths and Weaknesses for Major Design Alternatives

Alternative	Feasibility	Impact Estimator Bias	Estimator Precision	Cost
Experimental Design	Problematic for programs with excess capacity Ethical concerns likely May require	Spillover effects likely Best way to control for participant/non-participant differences	Likely to be constrained by small sample size	Likely to be least expensive alternative

	cooperation of referral sources			
Non-Experimental Design	Requires identification of reasonable comparison population Requires collection of data from comparison population	Unobserved differences between treatment and comparison groups may not be adequately controlled Outcome differences may reflect environmental differences	Less likely to be constrained by small sample size than experimental design	Data collection will be more expensive than under an experimental design, holding sample size constant, because it will come from two populations Obtaining the larger sample that this design makes possible will also add to cost
Random Outreach Design	Requires implementation of random outreach May require cooperation of referral sources	May be biased if program has different impact on participants induced by random outreach than on others Preserves use of randomization to control for participant/non-participant differences	Relies on effectiveness of random outreach More likely to be constrained by small sample size than experimental design	Requires larger sample than experimental design for given precision Outreach may be costly Analysis is somewhat more complex
Independent Evaluation of Multiple Sites	Evaluation may be tailored for each site=s program and circumstances Design and data collection constraints in one site need not constrain design in other sites	No problems other than as above	Will be very poor for small sites	Same comparison group may be used for multiple sites in non-experimental design
Joint Evaluation of Multiple Sites	Requires reasonable comparability of	Inappropriate pooling can cause bias, but this can be tested	Precision is substantially enhanced	Costs to resolve cross-site differences and

	target populations Common evaluation methodology for all sites may not be the best design for any single site		through pooling of data if target populations are sufficiently similar across sites	coordination requirements may be significant Economies of scale from multi-site evaluation will be realized Joint analysis of data is slightly more costly than separate analyses

Return to ToC

---

1. Differences in means include differences in percents for outcome variables that indicate whether or not an outcome for an individual satisfies a specific condition (e.g., has visited with the child at least once in the past week).

2. See Bloom, H.S. (1984). "Accounting for No-Shows in Experimental Evaluation Designs." Evaluation Review, vol. 8 (April), pp. 225-246. This simple formula relies on the assumption that participation has a constant effect on the outcome expected for an individual in the absence of participation, which may be incorrect. An equally simple formula is applicable under the assumption that the size of the impact for an individual is proportional to the individual's outcome in the absence of participation. See Chapter Eight for a discussion of other possibilities that allow for interactions between the magnitude of the impact and baseline characteristics of fathers.

3. As mentioned in a previous footnote, the magnitude of the program's impact may vary with baseline characteristics of the father. This issue could be conveniently studied in the context of the multivariate analysis.

4. This assumes a one-tailed test. See Exhibit VI.1.

5. Some program evaluations use non-participants from a program's target population, and/or program dropouts, for the comparison group. This approach is problematic because participants are self selected. Clever use of the data can sometimes solve the self-selection problem. See Bell, S. et al. (1995) Program Applicants as a Comparison Group in Evaluating Training Programs, Upjohn Institute: Kalamazoo, MI.

6. An earlier evaluation compared similar data for Racine and Fond du Lac Counties, the two pilot counties for Children's First. At the time (before 1991), the Racine program offered substantially more employment services than the Fond du Lac program, through JOBS, but not other substantial services. Measured impacts for the Racine program were substantially greater than for the Fond du Lac program.

7. See Goldberger, A.S. (1972) "Selection Bias in Evaluating Treatment Effects," Discussion Paper 123-72, Institute for Research on Poverty, University of Wisconsin-Madison.

8. As in the formula presented previously for the experimental design, this formula assumes that the impact of participation is the same for all participants. Interactions between participation impacts and baseline characteristics of participants can be incorporated in multivariate models.

9. For those familiar with multivariate selectivity models, the participation results would be used to construct an instrument for a dummy variable that identifies participants. The instrument's value would, in part, depend on the randomized outreach indicator and would be key to avoiding high collinearity between the instrument and control variables that might appear in the outcome

equation.

10. Even if there is no statistical advantage to joint evaluation of multiple sites, it may be economically efficient to have a single evaluator evaluate multiple sites simultaneously. There will be many common features of data collection instruments and other aspects of the evaluation, and experienced gained in implementing an evaluation of one site will benefit evaluations of other sites.

11. A process evaluation of each site would likely provide explanations for variation in program impacts across sites, although they would not be definitive.

CHAPTER FOUR

OUTCOME MEASUREMENT

I. Introduction

In designing and conducting an evaluation of responsible fatherhood interventions, the evaluator must first determine the primary program outcomes of interest. As fatherhood programs vary greatly, so do the outcomes these programs seek to achieve. Some programs may have only one or two primary outcomes, while others may address an array of factors related to responsible fatherhood.

The outcomes chosen for the evaluation should be those that are most directly related to the program goals and must be amenable to measurement. Some programs may already systematically document information on particular outcomes, which can serve as a starting point for determining those that should be included in a formal evaluation. A review of more than 300 fatherhood programs, however, found that the vast majority did not document outcomes in their programs.(1)

In the following section, we describe potential outcomes of fatherhood interventions, suggest specific measures that may be used in an evaluation, and discuss difficulties that may be encountered when developing measures for outcomes of fatherhood interventions.

II. Potential Outcomes and Methods of Measurement

Discussions with experts and examination of relevant literature yielded several potential outcomes for fatherhood interventions that may be categorized into five broad categories: (2)

- responsible behavior;
- father's relationship with child;
- father's support capabilities;
- child well-being; and
- the co-parental relationship.

The most common methods for measuring outcomes used in published studies are self-reports by the subjects (e.g. father, child, mother) and interviews conducted by trained staff. Other methods include observation and coding of behavior by a trained observer, and examination of public records (e.g. paternity status, employment, criminal activities); these activities, however, are done with less frequency given the significant resources necessary to conduct them.

There are several issues to be aware of when designing and using outcome measures for fatherhood programs. Subjective measurements, such as closeness and quality of the father/child relationship as measured by self-reports, are likely to differ depending on the person that is reporting the measure. For instance, a father might report feeling close to his child, but the child may report not feeling close at all. In addition, what the measures mean may differ across respondents, so that one child, for example, may view closeness very differently than another child. Deciding how to use responses from different groups, either separately or in combination, as well as standardizing responses from individuals are important issues to consider when designing outcome measures.

Matching outcome measures to program goals and characteristics is also crucial. Care should be taken to ensure that the outcomes to be measured are directly related to the actual goals of the program. In addition, how outcomes are measured can significantly affect how the results should be interpreted. For instance, having the mother, as opposed to the father, provide responses can greatly change the results.

It is also important to consider the time and resources necessary to measure the outcomes. For instance, having an expert observe and code interactions between father and mother may be a desirable way of measuring quality of co-parental interactions, but given resource constraints, it may not be feasible.

Below, we discuss a variety of outcomes that might be associated with fatherhood interventions and how these outcomes can be measured. We organize the discussion by the five broad categories of outcomes: responsible father behavior, father's relationship with child, father's financial capabilities/support, child well-being, and the co-parental relationship.

A. Responsible Behavior

Examples of outcomes that might be indicative of responsible father behavior include:

1. **Reduced Substance Abuse:** Whether the father uses/abuses drugs or alcohol, as reported by mother and/or father. Use and/or abuse may be defined in terms of frequency, quantities, and types of drugs or alcohol used, or in terms of the clinical criteria for a substance abuse diagnosis.(3)
2. **Reduced Criminal Involvement:** The nature and frequency of arrest and convictions of the father, as reported by father or as ascertained from public records.
3. **Reduced Unplanned Child-bearing:** Whether the father has subsequent children that are unplanned and/or out-of-wedlock, and whether they are with the same mother or different mothers, as reported by father and/or mother.
4. **Marriage/Stable Relationships:** Whether the father has married the mother, as reported by mother and/or father, married someone else, is involved in a stable relationship, or has reduced the number of sexual partners.
5. **Community Connectedness:** Whether the father participates in community activities or organizations (e.g. voting, church, philanthropy, community).
6. **Safe Sex Behavior:** Whether the father has knowledge of and practices safe sex behavior.

B. Father's Relationship With Child

Examples of outcomes that illustrate the nature of the father's relationship with his child include

1. **Contact/Visitation:** How often the father visits or has contact with the child and the duration of the visits, as reported by mother, father, and/or child.
2. **Paternity Status:** Whether the father has established paternity, as reported by father or mother, or as ascertained from government records.
3. **Type of Child-Related Activities in which the Father Participates:** A measure of how active the father is in his child's life, as reported by mother and father. This could include how regularly the father engages in activities such as providing child care, disciplining, dressing and grooming, moral training, running errands for and with the child, celebrating holidays/special occasions with the child, attending school/church activities with the child, engaging in recreational activities with the child, discussing the child's problems with the child, and taking the child on vacation.
4. **Parenting Skills:** Father's knowledge of child development; provision of prenatal and well-baby care and immunizations.
5. **Closeness:** A measure of how close a father and child are, as reported by father, mother, and/or child. For example, the child could be asked, "How close do you feel to your father?", and responses could be "not very close," "fairly close," "quite close," or "extremely close." A measure of closeness could

be taken from ratings on a variety of scales, such as a child's rating of "parental understanding," "trust," "respect," "fairness," and "affection." Alternatively, trained staff could be used to observe and record interactions between father and child.

C. Father's Support Capabilities

Outcomes that illustrate a father's ability to support himself and his child, financially and otherwise, include:

1. **Employment and Earnings:** Whether or not the father is employed and the level of his wages, earnings, and income, as reported by father.
2. **Education/Training Activities:** Whether or not the father has completed a given level of education, or is engaged in education and/or training activities (e.g., GED classes, enrolled in high school, college prep courses, vocational training).
3. **Child Support:** How much formal or informal child support and how regularly the father is paying, as reported by mother, father, and/or public records. Any other types of non-monetary support/services provided by the father. Whether the father has an understanding of and the ability to navigate the formal child support system.
4. **Other Responsibilities:** Does the father have a driver's license, library card, insurance, comply with local regulations/pay fines, etc.
5. **Work Ethic/Attitudes:** Father's attitudes toward work and relations with employers. Job duration and reasons for leaving employment may be indicators or work ethic and attitudes.
6. **Housing:** Whether the father has adequate housing and a permanent address.
7. **Physical Health:** Father's physical health and nutrition. Physical health can be measured as an overall rating (excellent, good, fair, poor) as reported by the father, and/or as the presence or absence of specific health conditions.
8. **Mental Health:** Father's mental and emotional health. There exist a number of scales available to measure depression and anxiety.
9. **Self-Awareness/Self-Esteem:** Father's level of self-awareness and esteem, engagement in self-development activities.
10. **Anger Management:** Father's ability to control anger and constructively address emotional problems.
11. **Ability to Deal with Racism:** Father's ability to cope with racism and racial discrimination.

D. Child Well-Being

Outcomes that reflect aspects of the child's well-being might include:

1. **Academic Achievement:** How well the child is doing in school, as reported by teacher, father, mother, and/or child. Alternatively, the child's performance can be measured through scores on achievement tests, which might be a more accurate measure.
2. **Social Behavior:** A measure of how the child interacts with others, as reported by the father, mother, and/or teacher. Measures could include types of behavior at home and in school, such as peer sociability, autonomy, aggression, attitudes towards strangers, obedience, leadership, self-confidence, cooperation, and communication skills. This measure could be difficult to interpret due to potentially biased reports from the father, mother, and teacher, as well as variation in how behavior is interpreted.
3. **Problem Behavior:** Type and frequency of delinquent behavior, as reported by the father, mother, and/or teacher. This could include such indicators as deliberate damage of school property, truancy, lying to parents about something important, taking something without paying for it, and injuring another person seriously enough to require a visit to a medical facility. Components of this measure could be difficult to interpret due to potentially biased reports from the father, mother, and teacher, as

well as variation in how behavior is interpreted.

4. **Child's Economic Status:** The level of household income and poverty status of the child.
5. **Safety in the Household:** The child's home environment, appropriate supervision, and personal safety.
6. **Physical Health:** Indicators of the child's physical health might include reports of the child's general health status by the mother, age appropriate level of development as measured by height and weight, days lost from school due to illness, whether the child has been immunized, and the child's access to health care and health insurance.
7. **Emotional/Mental Health:** A measure of how the child views himself/herself and the child's level of distress or dissatisfaction, based on responses to a self-esteem and/or a self-assessment questionnaire, reports by the mother, and/or whether the child has seen a psychiatrist or other professional about behavioral or emotional problems.

E. Co-parental or Team Relationship

Evaluators may be interested in determining the effect of a program on the relationship between a father and the mother(s) of his child(ren). Examples of outcomes that may reflect that relationship include:

1. **Agreement/Cooperation Concerning Child-Rearing:** A measure of how synchronous the mother's and father's views on child-rearing are, as reported by mother and father. Some components of this measure could include discussion of school problems and planning special events for the child.
2. **Father's Relationship with Child's Significant Others:** Father's relationship with mother's partner(s) and ability to deal with mother's attitudes towards his own partner(s). Also, his relationship with grandparents and other relatives of his child.
3. **Quantity and Quality of Communication Between Parents:** A measure of the parents' ability to communicate with each other about both parental and non-parental issues, as reported by mother and father. Alternatively, an expert could observe the interactions between the parents in order to more accurately assess the quality of communication; this, however, would require significant time and resources.
4. **Arrangement for Child Access:** Are there formal or informal arrangements for the father's access to the child and are the arrangements adhered to.
5. **Agreement on Child Support:** Whether the mother and father agree on the level of financial and non-financial child support provided by the father.
6. **Parents' Feelings Toward Each Other:** A measure of how parents feel about one another, as reported by mother and father. This could include such affect measures as guilt or anger toward the other parent, as well as measures of conflict between parents, including incidents of spousal abuse.

III. Summary

The outcomes and measures described in the preceding sections are only generic suggestions of possible fatherhood program outcomes. The actual set of outcomes and measures used in an evaluation will depend on the nature of the intervention being evaluated and the specific circumstances under which the evaluation is being conducted. It is unlikely that any particular program's impact evaluation would include all or even most of the outcomes described here.

The programs we visited varied somewhat in terms of the specific outcomes each program was designed to affect. For example, one program has a particular focus on reducing infant mortality and improving child health by increasing the involvement of the father in pre-natal and child health care. This is a very specific objective not shared by the other fatherhood programs we visited. Another program, through its arrangement with the county court system, has as one of its primary objectives, increasing the level and consistency of child support payments. This is only a secondary objective of the other programs we visited. There were, however, a number of outcomes the programs did have in common. These include (see also ***Exhibit 4.1***):

- increased education and employment;
- reduced alcohol and drug use;
- improved parenting skills;
- increased father involvement with his child(ren);
- improved attitudes or feelings toward children; and
- improved social and family interactions.

The above outcomes represent those that fatherhood *program managers* believed to be the most important outcomes that their programs attempt to impact. Through our conversations with government agencies and private funders we gained a sense of the outcomes that they, as funders, believed to be most important for fatherhood programs to address. From the *funder's* perspective, the most important outcomes include (see also ***Exhibit 4.1***):

- reduced unplanned child-bearing;
- reduced criminal involvement;
- increased paternity establishment;
- increased contact with child;
- increased formal or informal child support;
- increased employment and earnings;
- increased education or training;
- improved child behavior; and
- increased cooperation with mother concerning child-rearing.

Regardless of the outcomes chosen for inclusion in the evaluation, they should be ones that are directly related to the program's activities (i.e., there is a hypothesized relationship between program services and the outcome of interest) and they should be important and meaningful to the intended audience of the evaluation findings, whether that audience be program managers, funders, policymakers, or all of the above. Once the desired set of outcomes to be measured is established, the evaluator must develop survey questions to address each outcome.(4) We recommend the use of questions and measures from existing survey instruments to the greatest degree possible, especially if such instruments have proven validity.(5) The use of existing instruments and measures also facilitates the comparison of findings across studies.

Return to ToC

---

1. See Levine, Jim and Pitt, Ed (1995). New Expectations: Community Strategies for Responsible Fatherhood. Famil[illegible] and Work Institute. New York, NY.

2. See Appendix A for a list of experts with whom we have discussed fatherhood intervention outcomes.

3. See American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, for the criteria for a substance abuse diagnosis.

4. In this chapter, we have generally expressed the outcomes and measures as *levels* in the discussion. In some cases it may be

more appropriate to measure *the change* in outcome variables, rather than the level.

5. For a review of a wide variety of survey instruments designed to measure attitude and personality, see Robinson J.P. et al. (eds.) (1991), *Measures of Personality and Social Psychological Attitudes*. Academic Press, Inc. San Diego, CA.

CHAPTER FIVE

EXPLANATORY VARIABLES

I. Introduction

In this chapter we discuss the use of explanatory variables in an impact analysis. We begin with a discussion of the reasons for using explanatory variables in a multivariate analytical framework. We then describe explanatory variables likely to be included in an analysis of responsible fatherhood program outcomes.

II. Purpose of Explanatory Variables

While some program evaluations simply compare outcomes for treatment and control group subjects (e.g., difference in means and difference in percent analyses), more frequently multivariate techniques (e.g., multiple regression and logit), are used to compare outcomes after adjusting for a set of explanatory, or control, variables. There are several reasons for using explanatory variables in multivariate models, and an understanding of these reasons is helpful in determining the value of collecting data for explanatory variables in a specific evaluation, and the types of data to be collected.

The reasons for using explanatory variables in multivariate models are, in brief:

- To increase the precision of estimated program effects;
- To control for "confounding factors" in non-experimental designs that would otherwise result in biased estimates of program effects;
- To estimate interactions between individual characteristics (as captured by the explanatory variables) and program effects; and
- To generally improve our understanding of the determinants of responsible fatherhood program outcomes.

We elaborate on these reasons below.

A. Estimator Precision

Estimates of program effects based on a sample of outcomes for participants and non-participants are subject to random estimation error due to "idiosyncratic factors" -- factors other than program participation that affect outcomes for those individuals. The "standard error of the estimate" is the commonly used measure of how large estimation errors are likely to be. As a rule of thumb, the chance that the absolute value of the estimation error is greater than twice the standard error of the estimate is about five percent.

The size of standard errors depends on, among other things, how much variation there is in outcomes across fathers because of idiosyncratic factors. The more such variation, the more difficult it is to determine whether differences in treatment and control group outcomes are due to such factors rather than to program participation. Idiosyncratic variation can be reduced by including explanatory variables that explain some of that variation. For instance, if some of the idiosyncratic variation is due to variation in the age of the father, then using father's age as a control variable would explain part of the idiosyncratic variation. If the reduction in idiosyncratic variation is large enough, standard errors for estimates of program effects will fall.(1)

B. Controlling for Confounding Factors

If subjects are randomly assigned to treatment and control groups, then differences in outcomes between the two groups that are not due to the effect of the program are random. In the absence of random assignment, however, the differences may be due to systematic differences in the characteristics of subjects that are related to how the subjects were assigned to the two groups. For instance, if older fathers are more likely to be participants than younger fathers, and if age is positively related to a desirable outcome, then a positive difference between the mean outcomes for the treatment and comparison groups will be due, at least in part, to the fact that treatment fathers are, on average, older than comparison group fathers. Attributing this difference to the impact of the program would be misleading -- estimated program effects would likely overstate the true effects (positive bias). Using father age as an explanatory variable will remove this source of bias, as would controlling for other characteristics that may vary substantially across the two groups.

Many potentially systematic differences between treatment and control fathers are difficult to measure, for either conceptual or practical reasons. Those factors which remain constant over time can be controlled for by using a baseline (pre-program) value of the outcome variable as a control variable. For instance, the evaluators might compare the mean change in hours per week spent with the child over the evaluation period for the two groups, rather than mean hours at the end of the period.

Comparing changes in outcomes can be misleading, however, if the baseline value of the outcome variable is related to the individual's participation decision. For instance, suppose that fathers with low hours of child contact are more motivated to both increase hours and to participate in the program than those with more contact hours -- precisely because their current contact hours are low. Such fathers are likely to achieve greater increases in contact hours than those with higher initial hours even if they do not participate in the program, so attributing the full difference in the mean change in outcomes to participation will overstate the impact of participation on the outcome.

C. Program Interactions with Individual Characteristics

Not all fathers will respond to a program in the same way, and for policy purposes it may be useful to know that the program has more favorable effects on some classes of fathers than on others. Given limited resources, it may make sense to target benefits toward those fathers on whom the program is likely to have the most favorable impact.

The simplest approach to determining whether there are interactions between the impact of a program and father characteristics is to divide treatment and control or comparison group fathers into subgroups, based on the characteristics and to compare outcomes across treatment and control or comparison subgroups with the same characteristic(s). This approach will be unsatisfactory with small samples, however, as is likely to be the case for a responsible fatherhood program evaluation.

Given samples that are too small to make statistically meaningful treatment/control comparisons within subgroups, some success in measuring interactions may be achieved by specifying multivariate outcome models for all treatment and control or comparison group members in which dummy variables for program participation interact with explanatory variables for key individual characteristics -- characteristics that may be related to the size of the program's impact.(2)

D. Understanding the Determinants of Responsible Fatherhood Outcomes

While the main objective of an evaluation will be to determine the impacts of responsible fatherhood programs on the behavior of fathers and the well-being of their children, an evaluation can also enhance our general knowledge about the proximate causes of desirable, or undesirable, fatherhood behaviors. That is, the evaluation can help answer the question: What are the characteristics of fathers that are associated with the most desirable, or least desirable, outcomes? Such information could be useful in designing policies to promote responsible fatherhood, regardless of the program impacts.

III. Important Explanatory Variables and Their Measurement

The variables chosen for inclusion as explanatory variables in a multivariate model should be factors that vary across fathers in the sample and that are believed to influence or "explain" differences in the outcome being estimated. While the choice of explanatory variables will depend on the specific outcome being analyzed, the variables discussed below are likely to be important explanatory variables in an evaluation of responsible fatherhood program outcomes.

A. Demographic Variables

Demographic variables such as age and race/ethnicity allow the evaluator to describe the characteristics of fathers who participate in both the treatment and control groups, ensure that the two groups are comparable, and, if not, control for the differences by including demographic characteristics as explanatory variables in the multiple regression model. Demographic variables may also be important in explaining differences in the program outcomes of interest. Age may be measured as a single continuous variable representing years or as one or more categorical variables (e.g. age less than 18, 18 to 24, 24 and over). Race categories typically include black (African American), white (Caucasian), and "other." The choice of racial categories and whether or not to use race as an explanatory variable will depend on the race composition of program participants. In addition, ethnicity may be used as a control variable if there is reason to believe that there will be differences in outcomes between, for example, "Hispanics" and "non-Hispanics" or among subgroups of Hispanics.(3) Race and ethnicity may also be combined into a single set of variables.

B. Educational Attainment

Educational attainment, as measured at baseline, may be an important predictor of program outcomes. Educational attainment is most commonly measured as the highest grade or year of school completed. The variable may enter the analysis as a continuous variable (years of education), but more often is used as a categorical variable. An example of a categorical scheme for an educational attainment variable might be: less than high school education, high school graduate, and education beyond high school. A separate category for high school graduates with general educational degrees (GED) is often added.

Because fatherhood programs often serve very young fathers, it is important to devise educational categories that reflect "age appropriate" levels of education. For example, a sixteen year old who has not completed a high school education should not be grouped with a twenty-one year old without a high school education. As with all explanatory variables, it is important to choose categories that are meaningful in relation to the outcome being estimated and that contain more than just a few observations within each category.

C. Work History

Explanatory variables reflecting work history will be important to include in models that estimate program effects on work related outcomes such as employment and earnings. Factors such as years of experience and levels of prior wages or earnings are likely to be important predictors of post-program employment and earnings. Prior work experience may be measured as an indicator variable (has/has no prior experience in a formal job), as a continuous variable (number of years working in a formal job(s)), or as a categorical variable (no previous job experience, less than one year experience, 1 to 3 years, etc.). Prior earnings may be measured in terms of hourly wages and/or weekly/monthly/annual earnings. Depending on the outcome variable of interest, it may be important to know both components of earnings (hourly wage and hours of work), and therefore collect information on both wages and hours of work for use by themselves and to validate information collected on earnings.

D. Pre-Treatment Values of the Outcome Variables

The reason for including pre-treatment values of the outcome variables is that the evaluator will inevitably not be able to measure many of the factors that directly affect the post-treatment values of the same variables, and these unobserved factors are likely to have similar effects on the pre-treatment values. Including pre-treatment values of the outcome variables helps control for these unobserved factors. Very commonly, there will be multiple outcomes of interest, and a regression model should be estimated for each outcome.

Typically, only the pre-treatment value of each model's outcome variable is included among the explanatory variables for that model. Alternatively, one may use *the change* in the outcome variable as the dependent variable in the regression (as opposed to *the level*), omitting the pre-treatment value as an explanatory variable. This changes the interpretation of the regression estimates somewhat. For example, if affecting the level of child support payments is a program outcome of interest, the evaluator may estimate a model of child support payments using the post-program level as the dependent variable and the pre-program level as an explanatory variable. Alternatively, the effect of the program on child support payments may be estimated by using *the change* in the level of child support payments (the difference between pre- and post-program levels) as the dependent variable.

Inclusion of pre-treatment values of the outcome variable may substantially reduce the usefulness of other explanatory variables since the pre-treatment values of the outcome variable may capture most of the important effects of other variables on post-treatment outcomes. This, however, cannot be determined a priori and therefore it is important to obtain information on other explanatory variables. Further, these variables will be of interest for other reasons, such as analysis of impacts on specific subgroups and for use in participation analysis (discussed in Chapters Seven and Eight).

E. Site-Specific Factors

If conducting a multi-site evaluation, or if choosing a comparison group located in a different geographic area, it may be important to include variables reflecting environmental factors that affect the outcomes of interest and that vary by site. For example, if employment is one outcome of interest, it may be necessary to control for differences in labor markets across sites by including the unemployment rate as an explanatory variable. Another important environmental variable is the policy environment surrounding fatherhood related issues in a particular area. For example, child support enforcement methods and personnel in one area may be antagonistic toward fathers; in another area, they may operate in a manner that encourages cooperation with fathers. Other examples of environmental factors that may be related to outcomes of fatherhood programs include: the poverty rate, the rate of welfare recipiency, per-capita income, and crime rates.

Alternatively, a site-specific dummy variable may be used to capture all environmental differences across sites. This may be used if treatment and control groups within each site are believed to be comparable with respect to the important environmental factors, and therefore may be assigned the same site-specific dummy variable. For example, suppose a multi-site evaluation of a program operating in Cleveland and San Diego is conducted using a non-experimental comparison group design where the control groups for both sites are chosen from a geographic area adjacent to the area in which each program operates. A single site-specific dummy variable differentiating Cleveland from San Diego may be assigned, with the same value being assigned to both the treatment and control group within each site. This is possible only if it is believed that the environmental factors affecting the treatment group and control groups within each site are the same.

If, however, environmental factors differ between the treatment and control groups within each site (i.e., the adjacent geographic area from which the control groups were chosen differs substantially), then the site-specific dummy variable approach is not feasible. This situation would require assigning a different variable for each of the control and treatment groups at each site, but these variables would capture the treatment effect as well as the effects of differences in environmental factors.

F. Measures of Program Inputs

If program participants receive varied types and/or levels of services, or if identification of the impact of a specific service component on program outcomes is desired, then explanatory variables representing measures of program inputs should be included in the regression model. The measure may be expressed as an indicator variable, or as a continuous variable representing the number of "units" of a particular service component received (e.g., hours of case management, length of time spent in the program, number of parenting skills seminars attended, etc.).

Return to ToC

---

1. Adding explanatory variables that produce only small reductions in idiosyncratic variation may, however, result in larger standard errors. Each variable added uses up some of the scarce information in the sample (i.e., reduces the degrees of freedom), and collinearity among explanatory variables can increase standard errors.

2. A specification with interactions can be sufficiently general to be equivalent to separate analyses of subgroups. Hence, this strategy can only ameliorate the small sample problem if the specification is restrictive relative to separate subgroup analyses. For instance, impacts of other explanatory variables on outcomes may be assumed to be the same regardless of the value of the interacted explanatory variable(s).

3. Puerto Rican, Cuban, Mexican, Mexican-American, Chicano, and "Other Spanish" are widely cited Hispanic subgroups.

CHAPTER SIX

SAMPLING AND DATA COLLECTION

I. Introduction

In this chapter, we address issues related to the selection of the study sample and methods for collecting data on study participants. We begin with a discussion of methods for determining sample size and the process by which treatment and control/comparison groups may be selected. We then describe methods available to evaluators for collecting data on study participants, including surveys and program administrative data sources. We conclude the chapter with a discussion of the content and timing of baseline and follow-up data collection efforts.

II. Sampling Methodology

All of the evaluation design alternatives call for the identification of volunteer subjects for the study from one or more target populations. These volunteers will constitute the sample for the evaluation. The volunteers will be assigned by either a random or non-random methodology (depending on the design) to treatment and control or comparison groups. In this section we discuss issues related to identification of the target populations, recruitment of volunteers, assignment to treatment or control/comparison group, enrollment in the program, and the number of volunteers (i.e., the sample size) needed to obtain estimates of program impacts that have reasonable statistical precision (i.e., the sample size).

One issue that cuts across most of the issue areas considered below concerns the extent to which the program's "normal" process of recruitment and enrollment is maintained during the evaluation period. It seems inevitable that the process will be changed to some degree. Large changes, however, may make it difficult to generalize findings to fathers enrolled through the normal process. Hence, changes to the process that are made for purposes of the evaluation should be minimized, made in a way that is not likely to have an effect on the types of fathers enrolled in the program or the nature of the program itself during the evaluation period, and documented.

A. Defining the Target Population(s)

In the experimental and randomized outreach designs for a single program, the target population from which all study volunteers will be obtained will be the same for the treatment and control groups. In the non-experimental design, volunteers for the comparison group will come from a different, but similar, population as those for the treatment group -- the comparison target population. In a multi-site evaluation, volunteers will come from target populations at each site and, if a non-experimental design is used, from multiple comparison target populations. Below we first discuss issues related to the definition of the target population for an experimental or randomized outreach design, then consider issues concerning the selection of comparison target populations, and conclude with a discussion of the time frame for recruiting volunteers from the target population.

1. Target Population for Experimental or Randomized Outreach Design

The target population for obtaining study volunteers may be defined as the target population for the program being evaluated. The latter might be defined in many ways, such as all non-custodial fathers in some geographic area or as all non-custodial fathers who come in contact with the program's referral source(s). Using the program's target population as the target population for the study volunteers is important for assuring that the results of the evaluation are results for fathers with the characteristics that are "normally served" by the program.

There is, however, at least one important reason to consider going outside of the program's target population for the evaluation: to increase the number of volunteers by enough to ensure adequate numbers of treatment and control subjects, and to be sure that the program is not underutilized during the study period. If feasible, this should be done in a way that expands the population but does not materially alter the distribution of characteristics of fathers within the population. For instance, it might be possible to expand the program's target area into an adjacent area that has a similar population. Alternatively, if a program recruit's participants through a hospital maternity ward, fathers contacted and recruited through other hospital maternity wards could be added to the target population for the evaluation.

2. Comparison Target Population (Non-Experimental Design)

To the extent possible, the comparison target population should be matched to the treatment target population on characteristics of fathers and characteristics of the environment. Thus, if the treatment target population is all non-custodial fathers in a specific area, the comparison population would best be all non-custodial fathers in another area that have characteristics similar to those of non-custodial fathers in the treatment target population. The economic and policy environments of the two areas should also be similar. Alternatively, if the treatment target population is non-custodial fathers contacted through a hospital's maternity ward, the comparison target population could be non-custodial fathers contacted through the maternity ward of a similar hospital located in an area with a similar economic and policy environment.

In comparing the economic and policy environments in two areas, it is important to consider the possibility of differences in changes to the environments of the two areas over the evaluation period. For instance, if the economy improves in one area relative to the other, it will increase employment and perhaps child support among fathers in that area relative to those in the other area. One way to guard against this is to make sure the areas from which the two target populations are from are geographically adjacent and in the same local jurisdiction (e.g., county). The advantages of such proximity should, however, be weighed against the possibility of spillover effects -- interactions among the fathers in the two populations that might have an effect on the outcomes for either group.

It is likely that any comparison target population will differ in some respects from the program's target population. One way to increase uniformity would be to use a screening mechanism that screens out fathers with certain characteristics that are found in the comparison group population but not the target group population. For instance, if the target population only includes fathers from a specific minority population, then the screen would exclude fathers who are not in that same minority. Screens for age, place of residence, employment, and other factors might also be appropriate.

3. The Time Frame for Recruiting Volunteers

Volunteers for the evaluation will be recruited during a specified time frame. Given the small sizes of existing programs, it is tempting to have a long recruitment period to increase the sample size. An extended recruitment would obviously slow down the evaluation. In addition, the longer the recruitment period, the greater would be the risk that a change in the program or the environment during the recruitment and evaluation period would compromise the evaluation. Hence, the evaluators should be wary of using a very long recruitment period (more than, say, one year) as a means to increase the sample size. Extending the recruitment period is less of a problem in an experimental or randomized outreach design than in a non-experimental design because both treatment and control group volunteers would be subject to the same environmental changes.[1]

B. Recruiting Study Volunteers

The study subjects will be volunteers from one or more target populations. In this section we discuss issues concerning recruitment of the volunteers for the study.

We recommend that the evaluators identify and recruit volunteers for the study in the same way that the program normally identifies and recruits program participants. If the program advertises its services, then the evaluators would advertise for volunteers in the same general way. If an agency refers fathers to the program, then the agency, and perhaps similar agencies, should be asked to refer the same types of fathers to the evaluators.

When a potential volunteer is identified and contacted, the person should be told about the opportunity to participate in an "important research study about fathers who do not have custody of their children." If advertising is used to attract volunteers from the community, the ads should include some information about the study, including information on payments for volunteers who complete the study, and a toll-free number to call. If identified by a referring agency, agency staff should briefly explain the opportunity and offer similar information in written form. In this case it would be desirable to have the agency make a telephone available to the father for the purpose calling the evaluator. The father should also be told that volunteers will be paid a specified amount for completing an interview. The connection between the study and the program should not be mentioned, because volunteers who are not eligible to participate in the program might later be disappointed and upset. The role of agency staff should, in general, be kept to a minimum to avoid burdening them and to limit opportunities they might have to intentionally or unintentionally compromise the evaluation.

During the initial phone contact with the evaluator, the evaluator should:

- Explain the nature of the "study" more fully, focusing on its general purpose of improving the relationships between non-custodial fathers and their children and generally improving their lives (but not mentioning the program);
- Apply any screen that might be used to determine the caller's eligibility;
- Describe the baseline and follow-up interviews and the payments that the father will receive for completing them;
- Ask the father if he would like to volunteer immediately, and give the father the option of volunteering before a future date;
- Collect contact information if the caller is willing to provide it; and
- Make arrangements with those who volunteer for conducting the baseline interview.

It seems likely that many fathers who are initially identified as potential volunteers will not volunteer. Extensive efforts could be made to encourage volunteering, but they could ultimately be counterproductive because marginal volunteers might turn out to be very uncooperative study subjects and unlikely program participants. As described above, the process for obtaining volunteers allows fathers to "back-out," without embarrassment or other immediate consequences by simply not contacting, or re-contacting, the evaluator.

C. Random Assignment to Treatment and Control Groups, and Program Enrollment

Under the experimental and randomized outreach designs, volunteers from the program's target population will be randomly assigned to treatment or control groups. In this section we discuss the process of random assignment.

For the experimental design, we highly recommend that random assignment occur shortly following the baseline interview and also that the interviewer not be involved in the process. Knowledge of the opportunity to participate on the part of either the volunteer or the interviewer could have an effect on the quality and nature of the volunteer's answers, making answers for treatment group members less comparable to those for control group members. This could be accomplished by having the evaluator randomly assign the volunteer after being notified of the completion of the interview, but before reviewing the information obtained from the interview. When evaluation staff have information about a volunteer, they may be tempted to thwart random assignment so that especially "deserving" or "promising" subjects are assigned to treatment, or that undeserving or unpromising subjects are not. The process described above limits such opportunities.(2) Alternatively, study volunteers could be assigned to treatment and control groups based on their Social Security Number (SSN). For example, persons with SSNs ending with the numbers 0, 1, 3, 6, and 8 would be assigned to the treatment group, while those with SSNs ending with the numbers 2, 4, 5, 7, and 9 would be assigned to the control group.

When a volunteer is assigned to the treatment group, the evaluator should then take steps to implement the treatment. Under the experimental design, we recommend that the evaluator contact the program and give the program information needed to contact the volunteer. It would then be up to the program to recruit the father. Control group fathers would not be identified to the program, and would not be recruited.

It may be advisable to have the evaluator call the volunteers assigned to the treatment group before giving their contact information to the program to thank them for participating in the interview and ask them if they would like the opportunity to participate in a special program that helps non-custodial fathers and their children. Only fathers who reply affirmatively would be referred; others would presumably not participate in the program and would be part of the non-participant treatment group (see Chapter Three). If this is not done, some fathers who are contacted by the program following the interview may guess that the interviewer supplied their name to the program without the father's permission to do so.

A somewhat different process would be more appropriate for the randomized outreach design. In this case, the interviewer could ask the father if he wanted someone to contact him with information about a program to help non-custodial fathers and their children. All those who reply affirmatively would then be contacted by the evaluator's staff and provided with the information. The evaluator would also assign volunteers to treatment and control groups upon completion of the interview, but before examination of the interview data. The randomized outreach might be applied in one of two ways.

A simple way would be to have the evaluator provide information to program staff about treatment group volunteers, but not about control group volunteers. Then the program could conduct outreach activities to enroll the treatment group volunteers. A more costly and perhaps problematic way would be to have the evaluator conduct the outreach activities directly, with some outreach to control group volunteers and more intense outreach to treatment group volunteers. The latter method, including some outreach to control group cases, may yield more study participants. This would be important if the program would otherwise have excess capacity. The method has some distinct disadvantages, however, including: being more expensive; being susceptible to manipulation by evaluator staff; being different from the program's "normal" outreach efforts, and perhaps being less effective than comparable outreach that comes directly from the program. If increased program participation is desired, a better approach might be to offer more enrollment incentives to all volunteers upon completion of the baseline interview.

Under the non-experimental design, the only issue is enrolling treatment group volunteers in the program. As in the experimental design, the interviewer should not know to which group the father belongs.

D. Sample Size

One of the biggest challenges to an evaluation will be finding enough volunteers and eventual participants to obtain a reasonable level of statistical precision for impact estimates. In this section we discuss the relationship between the size of the sample, its division into participants and non-participants, and statistical precision.

The relationship between sample size and estimator precision depends on many factors that cannot be determined in advanced. This is especially the case when multivariate analyses are required to obtain impact estimates, as we would recommend (see Chapter Eight). Analysis of this relationship in the following much simpler situation is, however, indicative of what the actual relationship is likely to be.

Suppose we could simply randomly assign some fathers to participate in the program -- without the option of not participating -- and others to not participate. Then the treatment group would be synonymous with the participants and the control group would by synonymous with non-participants. Suppose also that the outcome of interest was a simple qualitative one; for example, has the father established paternity at the time of the follow-up interview? The difference between the percent of treatment and control fathers establishing paternity is an unbiased estimate of the impact of the program on establishment of paternity.

Even if the estimated difference in percent were positive (larger percent for the treatment), as we would expect, it might be positive just because, by chance, we happened to assign a larger share of fathers who would eventually establish paternity to the treatment group. To be confident that the difference was not just due to chance, we would require the estimated difference to be at least as large as a "critical value" --a value that has a small probability of being exceeded in a controlled experiment if the treatment does not have a positive effect. Formally, we would reject the null hypothesis of no impact (or possibly a negative impact) in favor of the alternative of a positive impact if the difference in percent is positive and larger than the critical value.

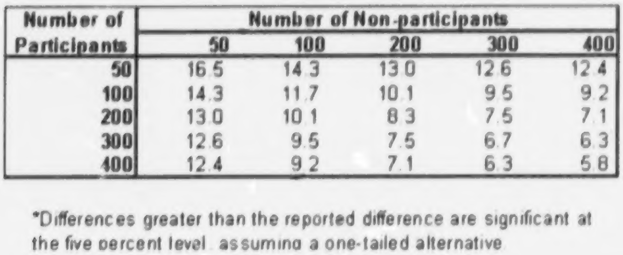
The critical value for this test depends on the sample sizes in the treatment and control group, the level of statistical significance desired, and the percent of fathers in the target population who do not establish paternity in the absence of program participation (the "population percent"). Increasing the sample size in either the control or treatment group reduces the critical value because it becomes less likely that a large difference is due to chance. The significance level is the chance that a difference will be greater than the critical value; choosing a smaller value requires the critical value to be larger. The population percent is unknown, but it can be shown that for given sample sizes and significance level, the critical value is greatest when the population percent is 50. In the absence of other information about this percent, 50 percent is often used to determine what the highest critical value would be for a given sample size and significance level.

Critical values for various sample sizes are shown in ***Exhibit 6.1*** for a five percent significance level (the most commonly used level) under the assumption that the population percent establishing paternity is 50. The difference in percent would have to be not only positive, but at least as large as the critical value to be significant at the five percent level (i.e., the chance of the difference exceeding the critical value if the program had no effect is just five percent). [3] In considering the numbers in the table, it is important to keep in mind that the sample sizes refer to volunteers who actually complete the study; numbers of initial volunteers needed to achieved a desired critical value may need to be substantially higher because of anticipated attrition.

Exhibit 6.1

Sample Size and Critical Percentage Point

Diffeences for a Difference in Percent Test *



Number of Participants	Number of Non-participants				
	50	100	200	300	400
50	16.5	14.3	13.0	12.6	12.4
100	14.3	11.7	10.1	9.5	9.2
200	13.0	10.1	8.3	7.5	7.1
300	12.6	9.5	7.5	6.7	6.3
400	12.4	9.2	7.1	6.3	5.8

*Differences greater than the reported difference are significant at the five percent level, assuming a one-tailed alternative

For most of the fatherhood programs we are familiar with, it would be difficult to obtain as many as 100 participants and an equal number of non-participants for the evaluation from a single program over a reasonably short period of time. For those sample sizes, the difference in percent would have to be as high as 11.7 percentage points to conclude that it is statistically significant. Is that large enough? This answer partly depends on how large the true difference would have to be for policymakers, funders, and others to conclude that it is an "important" difference. If a difference is not considered important unless it is at least 20 percentage points, then this sample would be of adequate size, but if a five percent difference is considered important it would not be.(4)

One program we visited had substantially more participants than the others. The Racine Goodwill Industries Fatherhood Program receives 55 to 65 court referrals per month, and enrollment for these fathers is mandatory. Presumably a six month evaluation enrollment period would yield over 300 participants. If, say, one or more other counties were used as comparison counties, it would presumably be feasible to generate 300 comparison cases. With this sample size, a difference of 6.7 percentage points would be statistically significant.

Before conducting an impact evaluation, it would be prudent to examine information that is indicative of the likely size of the program's effect on outcomes and compare that to the precision of the estimates for the likely sample sizes. As an example, for a non-experimental evaluation of the Racine program, cross-county differences in rates of compliance with court-ordered child support would be indicative of the possible size of the effect. If differences are only a few percentage points, significant effects are not likely to be found with samples of the size indicated. If, however, much larger differences are found, the evaluation may provide evidence that the program has a substantial impact.

The critical value can be lowered by increasing the size of either the treatment or control sample. Given a total sample size, the lowest critical value can be obtained by splitting the sample evenly between the two groups. In some situations it might be easier to increase the size of one group, but not the other, and there is no reason not to do this other than cost. For instance, if the program is filled to capacity, it would still be valuable to increase the sample size of the control group. In the randomized design it would be necessary to do this by changing the probability that a volunteer is assigned to treatment from 50 percent to some lower figure. This figure should be determined in advanced, based on anticipated numbers of volunteers and known program capacity; filling the program first, then putting additional volunteers into the control group, would violate random assignment.

Different sample sizes for the two groups might be preferred for other reasons, too. One such reason is ethical objections to assigning fathers to the control group when the program is operating below capacity; in this case, the share of volunteers randomly assigned to treatment could be increased above 50 percent, but at the cost of reducing the precision of the estimates.

III. Methods of Data Collection

The primary sources of data available to those who want evaluate responsible fatherhood interventions are surveys and program administrative data. We discuss issues associated with these two data sources in the sections below.

A. Surveys

Because most of the data necessary to conduct an evaluation of a fatherhood intervention will not be available from an existing source, the evaluation will necessarily rely on data collected through surveys of fathers, mothers, and, when feasible, children. The use of surveys facilitates the collection of uniform data across all study participants and allows the evaluator to collect information that otherwise might not be available.

In order to evaluate the impact of a program on specific outcomes, data on outcomes and other explanatory variables must be collected for both the treatment and comparison/control group before and after implementation of the program intervention. Therefore, at least two surveys will be administered to the study participants: a baseline survey and at least one follow-up survey. We discuss the timing and content of baseline and follow-up surveys in Section IV below.

There are a number of issues related to the gathering of information from a survey of participants and their families; most are not unique to the evaluation of fatherhood programs. These include:

- The design of valid questions and measures to capture the effects of the program;
- Whether self-administered surveys, telephone interviews, or in-person interviews should be used to collect information;
- Obtaining the cooperation of mothers and other family members; and
- The feasibility of collecting information from children.

These issues are likely to be resolved based on cost and feasibility considerations, and based on the outcomes of primary interest to the evaluators. We recommend, however, the use of in-person interviews for collecting the survey data. In-person interviews have several advantages over telephone or self-administered surveys. First, the response rate for in-person interviews is better than for self-administered interviews, and the number of incomplete answers are likely to be fewer. Second, the baseline and follow-up surveys required for a responsible fatherhood evaluation are rather lengthy. It may be difficult and uncomfortable to keep a respondent on the telephone for an extended period of time. Third, in-person interviews allow for the use of visual aids (e.g., flashcards listing potential responses) to illicit uniform responses. Finally, if there is a payment associated with the respondent's effort in completing the survey, that payment may be made directly to the respondent once the survey is completed.

An issue associated with the use of in-person interviews is where to conduct the interview. It may be more convenient for respondents to have the interviewer come to their home to administer the survey. Conducting an in-home interview, however, may have several problems: the area or environment may not be safe for the interviewer; there is a greater possibility of interruptions during the interview (e.g., the telephone, the presence of other family members); and the presence of other persons in the house who may overhear the interview may affect responses. For these reasons, it may be desirable to designate an easily accessible location where all interviews can be conducted.

Costs associated with conducting the survey will also depend on the outcomes of primary interest. For example, information on outcomes such as earnings, child support, paternity establishment, hours of contact with one's child, subsequent children out-of-wedlock, and drug or alcohol use may be easily collected from self-reports made by both fathers and mothers. If more complex information on child well-being, father well-being, father/child relationships, and father/mother relationships is desired, however, greater resources will be required both in developing and administering an instrument to measure these concepts.

Once the variables of interest have been determined, the evaluator must develop the survey instrument. We have discussed potential measures for program outcomes and explanatory variables in Chapters Four and Five. As discussed in those chapters, it is best to rely on survey questions that have been used in previous studies. Many national surveys collect information on many of the same variables that will be of interest to a fatherhood program evaluation. These instruments may serve as guides to the evaluator when developing questions for the baseline and follow-up surveys. Once the surveys have been developed, they should be "pre-tested" -- administered to a small sample of subjects -- to learn about possible problems (e.g., ambiguous questions, new alternatives missing from response lists), and correct them when possible. Pre-testing usually precedes from a slow "talk-through" with readily available respondents through interviews that are conducted with respondents from the target population as if they were participating in the actual survey. When the final instrument has been developed, the interviewers who will be administering the survey will need to be trained to ensure that the survey is administered correctly and uniformly across all respondents.

B. Administrative Data

Another potential source of information that may be used to conduct an evaluation is program administrative data. Most programs collect and maintain some information on their participants. One program we visited collects a variety of information on the initial application forms including:

- Demographic information: age, race, education, place of residence, living situation, marital status and other primary relationships, number of children and the ages, paternity/custody status, and AFDC participation status of each;
- Sources of income support, job training, skills, and interests;
- Criminal history, gun permit, and substance abuse information; and
- Expectations about what the applicant hopes to gain from the program.

In addition, this program is currently developing a follow-up database that will track outcomes for participants in the areas of paternity establishment, child support, arrears, visitation, employment, job duration, wages, educational attainment, and criminal activity. Follow-up information will be collected on former participants every six months.

These types of administrative data can be useful for conducting an impact evaluation, however, they suffer from one critical flaw: they are available only for persons who actually enroll in the program. Unless similar data can be obtained for the control or comparison group, a rigorous impact evaluation cannot be conducted using program administrative data alone. Data on outcomes for participants obtained through administrative can be useful for comparing the outcomes for participants at program completion to their outcomes as measured at follow-up (typically some months later). Such a comparison can illustrate temporary versus longer lasting program effects.

Another type of information often available through program administrative records which can be very useful to an evaluation is information on the types and levels of services that program participants receive.

It is important for an impact evaluation that all persons in the treatment group receive the same treatment. If analysis of information on program inputs reveals that participants receive different types or levels of services, then there may be reason to believe that estimates of program impacts will be biased. Unfortunately, this type of bias is difficult or impossible to control for since differences in services levels across participants is unlikely to be random. If there are differences in service levels/types across participants, it is probably because participants have different needs. The information and sample sizes necessary to correct for bias stemming from the selection between participants and services will not typically be available to evaluators.

IV. Baseline and Follow-Up Data Collection

In this section we discuss issues related to the administration of the baseline and follow-up surveys. We begin with a brief discussion of the content of baseline and follow-up surveys, and then discuss timing issues associated with survey data collection efforts.

A. Types of Data to Collect

The initial or baseline survey should collect information on all explanatory and outcome variables of interest. We refer the reader to Chapter Five for a discussion of potential explanatory variables for inclusion in the survey instrument, and Chapter Four for a discussion of potential outcome variables. The baseline survey will be more comprehensive than the follow-up survey because it is not necessary to collect follow-up information on characteristics that do not change over the observation period (e.g., date of birth, race, sex, source of referral, employment history, etc.). For purposes of the impact evaluation, follow-up surveys need only focus on collecting information on the outcomes of interest. In addition, follow-up surveys might include questions about whether study participants received any services similar to those provided by the program being evaluated from any other source. If treatment or control group members received services outside the program, it should be accounted for in the impact analysis.

Programs may wish to collect other types of information on a follow-up survey that may not be directly useful to the impact evaluation. For example, information on the participant's experience in program, such as which services he found most/least useful, can aid program staff in improving the effectiveness of their services.

B. Timing of Data Collection

Baseline Surveys: Ideally, the baseline survey should be conducted as soon after individuals are recruited into the study as possible, and before the individual has been assigned to the treatment or control group. This is for several reasons: First, it will ensure that interviewers do not know ("are blind to") the treatment status of the persons they are interviewing, and therefore will not introduce any unintended bias through the manner in which interviewers are administering the questionnaire. Second, it will ensure that interviewees' responses will not be influenced by referral to or subsequent contact with the program. Third, the more quickly the survey is implemented, the less likely individuals will be lost from the sample, either through subsequent lack of interest in participating in the study or because they can no longer be located.

Follow-Up Surveys: The length of time between conducting the baseline and follow-up surveys will depend on several factors: the length of time it takes a participant to complete the program; the particular outcomes of interest to the evaluation; and whether or not long-term impacts are of interest to the evaluation. In general, the follow-up survey should be conducted after a time interval that is sufficient for program participants to have completed the treatment, and for the treatment to have had an impact on the outcomes of interest.

Our site visits provided us with a contrasting example. One site has a defined six-week curriculum that has, as one primary focus, the goal of improving the employment prospects of young fathers. For this program, a follow-up survey may be conducted at a rather short interval following program completion, as the impact of the job readiness skills taught in the program are likely to have an immediate impact on the outcome of interest (employment).

In contrast, another site relies on intensive case management that focuses on changing the attitudes of participants so that they find in themselves the ability to achieve whatever goals they wish to pursue. For example, assume that employment is an outcome of interest to the program. The program treatment helps participants to realize that if they want to be employed in a decent job, they have the knowledge inside themselves to discover the means to do so. The treatment does not directly teach them job readiness skills, rather, the treatment induces them to go out and obtain the job readiness training or awakens the skills they already posses. A follow-up survey for participants in this program may need to be administered after a much longer interval because the treatment (learning self awareness and self empowerment) works indirectly on the outcomes of interest.

To conduct an evaluation of longer-term impacts of an intervention (for example, on paternity establishment, fathering of new children, interactions with children, educational attainment, employment, and substance abuse) one would want to obtain information on program participants for a three- to five-year period following participation (and possibly longer). Program administrators at the second program discussed above indicated that follow-up over a prolonged period could be a problem because participants are typically quite mobile and time-consuming to find. When outreach specialists were asked about tracking former participants for several years after participation, they indicated that for some it would be possible, but for others it would be very difficult. It should be noted that a previous study of this program met with mixed results in efforts to contact former participants.

Despite efforts to improve tracking procedures, attrition from the sample is still likely over a prolonged period, especially when tracking high risk populations. Differential attrition in participant and comparison groups could result in attrition bias in outcome comparisons. If follow-up interviews are to be conducted after long intervals, heavy emphasis must be placed on methods to reduce attrition.

Success in reaching study participants for shorter or longer-term follow-up can be enhanced by collecting more systematic contact information (e.g., information about friends and relatives) at the time of intake and at termination from the program. The offer of a monetary incentive for individuals responding to follow-up surveys or even to contact the program at various intervals in the future might enhance the ability of the program to track former participants over a prolonged period. Contact information as well as some useful follow-up outcome data may be obtainable through institutional records, though confidentiality requirements may be a difficult constraint to overcome. Unemployment insurance data, school records, criminal justice system data, and information maintained by welfare and child protective service agencies are examples of possible sources of information that might be used both to track participants and to collect independent outcome data. In addition, tracking fathers through their children, who may be easier to find, is another alternative. Finding that a father is not in touch with his child provides important information regarding certain outcomes of the fatherhood intervention.

Return to ToC

---

1. The evaluator should consider including control variables for date of recruitment in the multivariate models.

2. This problem is illustrated by the pilot test of SSA's Project NetWork demonstration. At the time, Lewin staff were helping design the baseline survey and we had an opportunity to review the pilot study data. We discovered systematic differences between the characteristics of "randomly assigned" treatment and control subjects. Upon investigation, it was determined that the case managers had influenced the assignment process to assign those with the best rehabilitation prospects to the treatment group -- a problem that was fixed for the later evaluation.

3. The test discussed here is a "one-tailed test" -- the null hypothesis of "no impact" is being tested against the alternative of a "positive impact." The null hypothesis is only rejected if the realized difference is positive and sufficiently large. Of course, the program could have a negative impact, in which case the realized difference is likely to be negative. Given the way the test is constructed, any negative difference, no matter how large, would lead to acceptance of the null hypothesis. This would be fine as long as the policy implications of a negative effect are the same as those for no effect. If they are different, the evaluator may want to use a two-tailed test. Use of a two-tailed test would increase each critical value in the table by almost 20 percent.

4. More precisely, if the true effect is an increase of 20 percentage points, then it is unlikely that this test would lead to a conclusion of "no difference," but if the true effect is only five percentage points we are likely to conclude that there is no difference. If effects as small as five percent are of little interest, then the conclusion of no difference in the latter case would have no serious consequences, while if a five percent difference is considered important, such a mistake would be unfortunate.

CHAPTER SEVEN

PARTICIPATION ANALYSIS

I. Introduction

In this chapter, we discuss important reasons why a participation analysis should be conducted in conjunction with an impact evaluation of fatherhood interventions, and present methods that may be used to perform such analyses. Participation analysis is usually an important component of formal evaluations of social interventions.

Participation analysis in its broadest sense concerns who participates in the program from among those who are in the program's target population. In an evaluation, however, participation analysis often focuses on who participates *conditional* on having participated in the study at some level. Under the three alternative designs we are considering, the participation analysis would focus on participation in the program conditional on volunteering to participate in the study. Under a randomized referral design, the analysis is conditioned further -- analysis of participation among volunteers who have been referred.

Analyzing who volunteers from among the program's entire target population is problematic because data on non-volunteers are not obtained in the baseline survey. Those who volunteer are not likely to be representative of all fathers in the program's target population; in particular, fathers who have the least desire to be responsible for their children are unlikely to volunteer.

Unless otherwise indicated, participation analysis, as used in the discussion below, refers to participation conditional on volunteering for the study.

II. Purposes of Participation Analysis

There are several reasons for conducting participation analysis in conjunction with an impact evaluation of a particular program. Below, we describe the three reasons we believe to be most relevant to the evaluation of fatherhood interventions. These include: increasing knowledge of the determinants of program participation; controlling for selection bias; and assessing the effectiveness of outreach and recruiting activities.

A. Increase Understanding of the Determinants of Participation

The information obtained from conducting a participation analysis can help program staff, funders, and policymakers develop a better understanding of the factors that determine the likelihood that fathers will participate in the program. This can be useful for a variety of reasons. Improved knowledge of the characteristics of those who participate may allow program staff to better tailor their services to those who are demanding them. Participation analyses may also help identify factors that inhibit fathers from participating, allowing program staff to address such potential obstacles to participation. Finally, participation analysis provides the information needed to estimate program participation for populations not previously served by the intervention.

B. Control for Selection Effects

The three evaluation designs rely on the comparison of outcomes for participants to those of non-participants. As discussed in Chapter Five, explanatory variables may be used to control for observable differences between the two groups when estimating the impact of an intervention. If, however, there are unobserved differences between the two groups that are systematically related both to participation in the program and the outcome of interest, the estimate of the treatment effect will be biased. This type of bias is referred to as selection bias. Here, we discuss three potential sources of selection bias: self selection, program selection, and attrition.

Self Selection: Bias can arise if fathers who are more likely to succeed or have positive outcomes are also those most likely to participate in the program. For example, unobserved characteristics such as self-discipline and motivation may affect an father's likelihood of participating in a fatherhood intervention. These same characteristics may also positively affect many of the outcomes of fatherhood interventions, such as contact with the child, employment, and child support. If fathers with these unobserved characteristics are more likely to participate in a fatherhood program (i.e. are *self selecting* into the program) then estimates of the impact of the program may be biased. In this example, the estimated program impact would be greater than the true impact. Those who participate would have more contact with their children, be more likely to be employed, and pay more child support relative to those in the comparison group, even in the absence of the program.

Program Selection: Bias resulting from program selection effects may occur if program referral, recruiting, or acceptance policies systematically screen-out particular types of fathers from the program. If screening criteria used by program staff are related to the outcomes of interest, there is the potential for selection bias. For example, the Indianapolis FRP uses a rather intensive pre-screening application process. The pre-screen involves several interviews with FRP staff to inform the applicant about what the program involves, to determine how serious the applicant is about participating, and to assess the applicant's ability to participate and potential for successful completion of the program curriculum. The purpose of the pre-screening is to identify and enroll those most likely to succeed in the program. If this manner of participant screening is not accounted for in the evaluation design, the estimated program impact will be biased upward.

Attrition: If participants who drop-out of the program before completion have unobserved characteristics that are systematically related to program outcomes, attrition bias may result. Such a situation is analogous to the self selection bias example described above. In this case, participants are self selecting *out* of the program. If, for example, participants who drop out of the program are less motivated or less willing to work than those who remain, estimates of the program effect on outcomes such as employment and earnings may be greater than the true impact.

Attrition bias may also arise if follow-up data on some comparison group members cannot be obtained, and these fathers are systematically different from those for whom follow-up data is available. For example, comparison group members who cannot be reached for follow-up may be persons without a stable residence, no telephone, or who become incarcerated or institutionalized. These characteristics are also likely to affect employment and earnings outcomes. In this example, estimated program impacts will be smaller than the true impacts.

C. Assess the Effectiveness of Outreach/Recruiting Activities

Participation analysis may also be used to assess the effectiveness of outreach and recruiting activities. Participation analyses can estimate the effect of specific outreach or recruiting activities on the likelihood that fathers will become program participants. The opportunity to perform a rigorous assessment is offered by the randomized outreach design, where the individuals who receive the outreach (and the type of outreach they receive) are randomly selected, and therefore, selection bias associated with outreach methods may be minimized. Participation analysis may also be used to determine whether the outreach and recruiting activities undertaken by the program are attracting participants from the intended target population.

In general, the assessment of outreach and recruiting activities are restricted to outreach methods which are targeted to specific individuals or groups, rather than aimed at all individuals in a particular area (e.g., ad campaigns at specific locations like schools or churches, versus radio ads that reach an entire area). This is because it will not be easy to determine who has and has not received the outreach when broad-based methods are used. Analysis of the effectiveness of outreach efforts will also be limited by the fact that only the effect on study volunteers can be determined.

III. Conducting Participation Analysis

In this section, we present an approach to conducting participation analyses. We begin with a discussion of problems associated with defining and measuring program participation. We then present the steps to conducting participation analyses: computing sample descriptive statistics, and conducting multivariate analyses. A more technical description of multivariate participation analysis appears in Appendix E.

A. Measuring Participation

One of the more complex aspects of evaluating responsible fatherhood programs (and most initiatives targeting at-risk youth and adults) is determining who is actually being served. Most of the programs we visited (as well as the literature on responsible fatherhood programs) emphasize the importance of providing services that are client-driven and flexible. As a result, potential participants may have several contacts with the intervention before formally enrolling in the program, and some participants may never be formally enrolled. Even after the commitment is made, the participant may come, disappear for a while, and then return for services. While this flexibility may be essential to a successful program and working with an at-risk population, it can complicate the evaluation process because it makes it difficult to determine when someone has become a participant and, in some cases, stopped being a participant.

Evaluation researchers generally identify two basic approaches for defining participation in programs: (1) whether an individual has completed the formalized intake process (e.g., completed an intake form); or (2) exit or completion status.[1] As discussed below, both approaches have potential drawbacks when applied to responsible fatherhood programs.

There are two main problems associated with using a formal intake process to determine program participation: (a) some fathers who complete the formal intake process may subsequently receive few services, or (b) some individuals (e.g., related family members) who do not complete the formal intake process, may receive program services. The completion of a formal intake form may or may not reflect actual and full participation in the program. For example, in the IRFFR program, it is possible for fathers to attend some and even many group sessions without completing the formal intake process. Those attending these group sessions are only asked to record their names as attending the sessions. Formal intake into the IRFFR program in Cleveland occurs when an individual is assigned to an outreach specialist and completes an intake interview. During the initial home visit, an outreach specialist interviews the individual (and perhaps other family members) and completes the intake form. At this point, the individual becomes a protégé and is expected to be available for home visits by the outreach specialist and to attend group counseling sessions. Typically, the outreach specialist would continue to meet several times a month with the individual (and perhaps other family members) to discuss and monitor goal achievement over a three- to six-month (or longer) period. Because of the tailoring of the intervention to each protégé's needs and desires, the duration of participation and types of assistance received varies considerably across protégés.

Completion of the formal intake process does not necessarily mean that the individual completes the program or even moves much beyond completion of an annual plan. For example, in one of the case records that we reviewed, a father was visited several times by an outreach specialist, completed the intake form, and disappeared from the program shortly thereafter (indicating that he expected to return in a month or so). Hence, the inclusion of fathers based on completion of the formal intake process may result in the inclusion of individuals who subsequently receive few or no services -- and hence, may result in the inclusion of individuals into the participant group who have received few, if any, substantive services.

Another potential problem with using formalized intake to determine participation is that there may be fathers and family members who do not complete the intake process, but nonetheless receive services either directly or indirectly through the program. For example, in one of the programs we visited, family members of protégés or individuals for whom funding for the outreach specialist's services could not be obtained often do not go through the formal intake process but may participate in counseling and group sessions and may be greatly affected by services delivered through the program. An alternative may be to define participation as "receipt of at least one service" as opposed to "completed formal intake".

The second approach discussed in the literature -- defining participation in terms of exit from the program or completion status -- has its own set of problems. First, it is sometimes difficult to determine when a participant has completed or is "exiting" a program. As discussed earlier, programs for at-risk youth and adults typically are flexible in terms of service provision (e.g., one site uses an outreach specialist to tailor service delivery to the client's specific needs) and may not impose penalties for irregular program participation. Of the five programs we visited, only two had rather strict participation requirements. Thus, while there may be a core of services that participants generally receive (e.g., in-home counseling and group sessions), it can be difficult to define a core set of program activities that must be received before the individual is considered to have completed participation in the program.

Second, even if a core set of program activities can be defined, individuals who do not receive this core set of activities but receive a substantial level of activities will not be included as participants -- and the evaluation will miss important potential impacts of the program. Programs serving high risk fathers often encounter high rates of attrition, though the administrators at one site indicated that attrition was not a problem for their program. If "participation" is based upon completion of a core set of activities, the evaluation could miss a significant number of individuals who received some (or perhaps a considerable) level of services.(2)

For many programs it may be very apparent what constitutes a "participant". One program we visited has a very structured program with uniform services provided to all participants over a defined, and relatively short, period of time. In this case, it is very easy to determine who is and is not a participant --the father is either attending the daily classes or he is not. For most of the other programs we visited, this was not the case. Staff at two of the programs we visited indicated that they would have difficulty determining exactly how many active participants they currently serve due to the irregular participation of many of the fathers.

In evaluating fatherhood programs, careful attention should be given to the definition of what constitutes program participation. It is possible (and probably likely) that definitions will vary across responsible fatherhood programs according to the targeted population, and the structure and types of services provided.

B. Sample Descriptive Statistics

A first step in conducting participation analysis is to tabulate sample descriptive statistics on:(1) answers to survey questions concerning fathers' knowledge of the program and why they did or did not participate; and (2) characteristics of fathers that are thought to have an influence on participation. A comparison of descriptive statistics for these factors across participants and non-participants can identify factors that are important in determining participation.(3) Descriptive statistics also provide an overview or profile of fathers participating in the study.

In ***Exhibit 7.1***, we illustrate how sample descriptive statistics may be compared for purposes of the participation analysis.(4) For this example, assume that a non-experimental evaluation design is used. In comparing the means across the treatment and comparison groups, we subdivided the treatment group into those who actually participated in the program and those who chose not to participate. This latter group, the treatment group non-participants, can be further subdivided to differentiate between those who chose not to begin the program (no shows) from those who began but subsequently dropped out of the program (drop outs). If attrition in the control group is a problem (i.e., follow-up data for many individuals in the control group could not be obtained), then subdividing the control group into those with and without follow-up data may be necessary.

The structure of the table will depend upon the evaluation design selected. We have assumed that comparison group fathers are unable to enroll in the program and would have no reasons to be aware of its existence. In an experimental design, control group fathers may be aware of the program, but not be allowed to participated. The evaluator may find it useful to ask control group fathers what they knew about the program, whether they would have liked to participate, and whether they sought assistance from other sources because they could not participate in the program. In the randomized outreach design, the control group would also be subdivided into participants and non-participants.

A preliminary comparison of subgroup means may identify potentially problematic differences between participants and non-participants. A simple comparison of means, however, will not illustrate whether the differences are important enough (i.e. statistically significant controlling for all factors) to warrant the use of statistical methods to correct for potential selection bias in the estimation of the treatment effect. In order to determine the significance of the differences between participants and non-participants, and to control for these differences using statistical techniques, a multivariate analysis is necessary.

C. Multivariate Analysis of Participation

The details of the participation analysis will depend on which type of evaluation design is used (experimental, non-experimental, or randomized outreach) and on whether a single-site or multi-site evaluation is performed. We begin by discussing an approach to participation analysis for an experimental, single-site evaluation, then consider modifications necessary for the alternative designs and for a multi-site evaluation.

1. Participation Analysis under an Experimental Design

Under an experimental design, randomly selected volunteers are referred to the program (the treatment group) while others are not (the control group). We assume that control group members do not have the option of participating -- an assumption that is relaxed in the randomized outreach design. Hence, only the volunteers who are assigned to the treatment group can choose whether or not to participate.

The evaluator will need to estimate a particular type of multivariate econometric model for the participation decision -- a "binomial choice" model. The "logit" and "probit" models are the two most commonly used models in this general class. Such models specify that the probability of participation for an individual is a function of a set of explanatory variables. These variables should include baseline variables thought to have an impact on a father's participation decision. It should also include variables that might be used by program staff to decide whether to include a father in the program. These are the same variables that would be used to descriptively compare participants to non-participants within the treatment group, as well as to compare control group fathers to treatment group fathers.

The estimated model can be used to calculate the change in the probability of participation associated with a change in each explanatory variable. For instance, the evaluator may be able to calculate how the probability of participation increases or decreases with the age of the father, the age of the child, the current employment status of the father, etc.

The success of the participation analysis in identifying factors that are significantly associated with the likelihood of participation will depend on both the total sample size for the treatment group and the split of the treatment group into participant and non-participant subgroups. If the sample is small, it may be that all sample fathers with some specific characteristic (e.g., fathers under the age of 18) will all be in either the participant or non-participant subgroup, in which case it will not be possible to investigate the effect of that characteristic on participation other than to acknowledge that, in the sample, that characteristic alone is a perfect predictor of participation.(5)

The estimated model can also be used to compute a "conditional participation probability" for each father in the treatment group -- the probability that the father participates given his observed characteristics alone. The probability is conditional in the sense that it doesn't take into account unobserved factors that affect the father's actual participation decision. It answers the question: "What proportion of fathers with the same characteristics would participate if faced with the same decision?" The estimates reflect the fact that the father was referred to the program and also reflect any screening criteria that are applied by the program in accepting fathers.

Conditional participation probabilities have two specific uses. They can be helpful to a start-up program that is trying predict "demand" for its services, presuming the program has some information about the characteristics of fathers in the target population. A less obvious, but perhaps more important, use is in the impact analysis. As will be discussed in Chapter Eight, conditional participation probabilities play a critical role in separating the impact of a program from "selection effects" -- the effects of self-selection by fathers and screening by programs on differences in outcomes for participants and non-participants.

2. Participation Analysis in a Non-Experimental Design

The appropriate methodology for participation analysis in a non-experimental design is the same as for the experimental design. As we have described that design (Chapter Three), volunteers in the treatment group are in that group rather than the comparison group for reasons that are beyond their immediate control (e.g., their area of residence, or the hospital in which their child was born). Hence, the only choice they have is whether or not to participate in the program when offered the opportunity. This is no different than the choice offered to treatment group fathers under the experimental design.

While the methodology is the same under the two designs, the results have a different interpretation. In the experimental design, results are for volunteer fathers who have been referred to the program, while in the non-experimental design they are for volunteer fathers who happen to be in the program's target treatment population. Thus, the results are conditional on different recruitment and, perhaps, screening mechanisms. These mechanisms must be kept in mind when interpreting the findings.

3. Participation Analysis under a Randomized Outreach Design

In the randomized outreach design (Chapter Three), study volunteers are randomly assigned to receive strong (treatment) or weak (control) outreach. Fathers in either group may decide to participate in the program, but the differences in outreach are expected to result in higher participation rates among fathers who receive the treatment outreach.

Under this design, data for both the treatment and control groups would be used in the participation analysis because fathers in both groups choose whether or not to participate. One of the variables to be included in explanatory variables for the multivariate analysis would be an indicator for the treatment outreach. If the evaluators use multiple types of randomized outreach, the explanatory variables would include indicators for all types. They might also include variables measures of outreach intensity (e.g., the size of any monetary incentives). The coefficients of the treatment outreach variables would measure the impact of the variables on the propensity to participate, and could be easily converted to estimates of the effect of outreach on the probability of participation.

It is likely that participation analysis will be more fruitful under the randomized outreach design than under the experimental or non-experimental designs, for two reasons. First, holding the number of volunteers for the entire study constant, the number used in the participation analysis will be much higher under the former design than under either of the latter -- twice as large if subjects split equally between treatment and control groups. Second, the effects of the randomized outreach itself can be rigorously studied under this design, and may yield results that are important to both program operators and policy makers.

The evaluator may also find it useful to examine whether the effectiveness of the demonstration outreach varies with characteristics of fathers. For instance, the outreach may have been more effective for fathers in some age groups than in others. This will be feasible if the sample size for the evaluation is sufficiently large.

4. Participation Analysis in a Multi-site Evaluation

Opportunities for conducting informative participation analyses are improved in a multi-site evaluation beyond the opportunities available from independent evaluations of each site because of the possibility of pooling data from two or more sites. This will be especially important if sample sizes at individual sites are too small to support meaningful participation analysis.

The same multivariate methodology would be applied in a pooled analysis, but the explanatory variables need to be modified appropriately. Most importantly, variables to indicate the site should be included because participation is likely to be higher in some sites than in others even after controlling for observed baseline characteristics of individual fathers. Cross-site differences may be due to unmeasured environmental differences across sites, unmeasured differences in the target population, and/or unmeasured differences in program administration and the appeal of the program to potential participants. Another possibility is to allow for different effects of various factors across sites. In the extreme, this could mean estimating separate models for each site, but this would result in the loss of any advantage that might be gained from pooling the data. Because sample sizes for each site are likely to be modest, it would be prudent to pool the data unless there are strong prior reasons to believe that the effects of the explanatory variables on participation vary across sites.

The participation analysis for a multi-site evaluation under a randomized outreach design should also include dummy variables to indicate the site. These would capture the effects of all site-specific factors that have an impact on participation at the each site -- unique features of the environment, the program itself, and the target population. In addition, the evaluators may want to interact site dummies with the outreach treatment dummy or, if applicable, the multiple outreach variables. This would allow the evaluator to test the null hypothesis that the effect of the randomized outreach on participation is the same at all sites, and to estimate differences in effects across sites. Such an analysis might be helpful in providing information about subtleties of outreach, or about the environment in which outreach is conducted, that increase or reduce its effectiveness, especially when conducted in conjunction with a process evaluation.

Return to ToC

---

1. See Martha Burt and Gary Resnick (1992). "Youth At Risk: Evaluation Issues," prepared for the U.S. Department of Health and Human Services.

2. If an experimental design is used in the evaluation, all fathers assigned to the treatment group would be analyzed and a correction for "no shows" would be employed.

3. Comparison of means includes comparison of percents for categorical variables.

4. The variables indicated in the exhibit are for illustrative purposes only.

5. Including an indicator for the characteristic in Z would result in a computational failure in maximizing the likelihood function because increasing the magnitude of the coefficient in one direction would always increase the value of the likelihood function.

CHAPTER EIGHT

IMPACT ANALYSIS

I. Introduction

In this chapter we discuss the analyses of the evaluation data that will be necessary to estimate the impacts of responsible fatherhood programs. Given the preliminary and general nature of this evaluation design, the analysis methods discussed are intended to be illustrative of the methods that will be required. We provide a non-technical discussion of the methodology here; a technical presentation of the methodology appears in Appendix E.

In general, the impact evaluation will examine differences between outcomes for participants and non-participants. While the easiest way to conduct such an analysis is to compare differences in means or percentages of outcome variables for the two groups, outcome differences may reflect factors other than the impact of the program -- especially systematic differences due to the selection of study volunteers into the participant and non-participant groups, as well as others. We recommend using more complex multivariate methods in order to address these issues.

The selection issue we focus on in this chapter is the selection of study volunteers into participant and non-participant groups. As discussed in Chapter Three, participants and treatment group subjects are not synonymous. In all three designs (experimental, non-experimental, and randomized outreach) some treatment group subjects will choose not to participate, and in the randomized outreach design some control group subjects will participate. The methodology must, then, explicitly recognize the difference between "treatment" and "participation."

There are two other selection issues that we do not consider, but that deserve mention. The first is self-selection of study volunteers from the target population. Outcomes for study volunteers are likely to differ systematically from outcomes for other fathers in the target population, regardless of participation, and the impacts of participation on study volunteers may also differ from those that might be achieved for non-volunteers were they to participate. Studying the selection of volunteers would provide information about the extent to which estimated impacts of the demonstration would generalize to other fathers, but such a study would be difficult and costly to perform. We recommend, instead, that scarce resources be used to obtain estimates of the impacts of participation on those who volunteer.

The second selection issue that we will not consider further is attrition of study volunteers. No matter how intense the effort to obtain follow-up data from all volunteers, some will inevitably be lost to the study. This issue would be essentially the same as the issue of self-selection of volunteers if attrition were unrelated to program participation; then, those who leave the sample could be viewed as non-volunteers. It is quite possible, however, that attrition will be related to program participation, with participants less likely to drop out than non-participants.(1) Further, attrition among participants could be related to outcomes, with fathers who have less favorable outcomes more likely to leave the sample. If attrition rates vary substantially across participants and non-participants, then some effort should be made to correct for possible attrition bias.(2)

We focus on the analysis of data collected under an experimental study design (see Chapter Three), but also discuss how the analysis would need to be modified under each of the two alternative designs (non-experimental and randomized outreach). Differences in methodologies for the three alternative designs are subtle, but important.

We first present a methodology for evaluating the impact of a single program at a single site (Section II). To simplify the presentation, we discuss how the analysis would proceed for a single outcome variable that is assumed to be a continuous variable with an unlimited range (e.g., a child's score on a psychological assessment of anxiety or depression). This model can be repeated for multiple continuous outcome variables. We also discuss extensions of the model to qualitative (e.g., paternity establishment) or limited dependent variables (e.g., level of child support payments).

After completing the discussion of the methodology for evaluating the impacts of a single program at a single site, we discuss a methodology for jointly evaluating the impacts of multiple programs and/or multiple sites of the same program (Section III).

II. Methods for Analyzing Program Impacts at a Single Site

A. Analysis for a Continuous Outcome Variable under an Experimental Design

1. Difference in Means Analysis

In an experimental evaluation, fathers are randomly selected for referral to the program. If all treatment fathers participate in the program and all control fathers did not, then the impact of the program on a continuous outcome variable could be measured as the difference in means for the treatment and control groups. If the sample sizes are reasonably large, random assignment to the two groups makes it very likely that any substantial difference in means is due to the program and not due to random differences in the characteristics of fathers in the two groups, which are likely to be small.

Some fathers who are referred do not, however, participate in the program.(3) Presumably their outcomes would be more favorable if they did participate. If so, the difference in mean outcomes is likely to understate the program's impact on the average eligible father. The difference in mean outcomes will still be an unbiased estimate of the impact of *referring* fathers to the program, but funders and others are more likely to be interested in the impact on fathers who actually participate because it is only those fathers that make use of substantial program resources.(4)

One might be tempted, instead, to use the difference in mean outcomes for participants (i.e., for the subset of treatment group fathers who participated) and non-participants (i.e., for control group fathers plus non-participating treatment group fathers). This is likely to overstate the program's impact because participating fathers may be more motivated than non-participating fathers, and thus may achieve better outcomes than non-participating fathers even without participating in the program.

Instead, an unbiased estimate can be obtained by dividing the difference in mean outcomes for the treatment and control groups by the share of the treatment group that participates, as described in Chapter Three. This corrects for the fact that only a share of the fathers in the treatment group actually participate in the program. Because the share who participate is less than one, the resulting estimate will be larger than the difference in the mean outcomes for the treatment and control groups.(5)

2. Regression Analysis

For reasons discussed in Chapter Five, the evaluator may want to control for the influence of other explanatory variables on the outcome variable in estimating the effect of participation. If all treatment fathers participate, this is accomplished through a regression analysis of the outcome variable. The regression model specifies that the outcome variable is a (linear) function of a set of explanatory variables, and the coefficient of each explanatory variable represents the effect of a change in that variable on the expected value of the outcome variable, holding all other explanatory variables constant. One of the explanatory variables would be a dummy variable to indicate whether the individual is in the treatment group; other explanatory variables would represent baseline characteristics thought to have an effect on the outcome variable. The estimated coefficient of the treatment dummy would be the estimate of the treatment effect.

If all referred fathers did participate, the expected value for this estimate of the program's impact on the outcome variable would be identical to the expected value of the difference in mean outcomes for the treatment and control group. The only reason we would prefer this estimate to the difference in mean estimate is that the standard error of the estimate would be lower -- provided that we had judiciously chosen as the other explanatory variables a set of variables that explained substantial amount of the variation in outcomes across fathers.(6)

All referred fathers will not participate, however, and the coefficient of the treatment dummy from the regression will be too small (biased toward zero) as an estimate of the program's impact on participants, just as the difference in mean outcomes for treatment and control fathers would be. We could, instead, replace the treatment dummy with a participation dummy, which would be coded as one for treatment fathers who participate only, and zero for everyone else. The estimated coefficient of the participation dummy would likely be too large as an estimate of the impact of participation (biased away from zero) for the same reason that the difference between the mean outcomes for participants and non-participants is too large: those fathers who choose participate are likely to be more highly motivated and have better outcomes than those who choose not to participate even in the absence of participation.

The solution to the bias problem in the regression approach is a mathematical extension of the solution used in the difference in means approach, although not obviously so. Instead of using either the participation dummy or the treatment dummy as an explanatory variable in the regression model, the analyst should use a "modified participation variable" that, like these two variables, is zero for all control group fathers, but that is equal to an estimated "participation probability" for treatment group fathers. Specifically, the value assigned to treatment group fathers, whether or not they actually participate in the program, should be the estimated conditional participation probability obtained from the participation analysis (Chapter Seven and Appendix E). Use of this value instead of a value of one for all treatment group fathers is analogous to dividing the difference in mean outcomes for the treatment and control fathers by the share of treatment fathers who participated in the program.

While it is relatively easy to obtain this "two-step" regression estimator of the participation effect, computation of standard errors is more problematic because correct standard errors need to take account of estimation errors in the participation probabilities. Further, use of a maximum likelihood estimator for the *joint* participation and outcome models, or some other joint estimator that is computationally simpler, may produce more estimates with lower standard errors.

Two features of the regression methodology deserve further attention before we turn to variants for alternative evaluation designs. First, the methodology can be used to estimate participation effects even if there is no control group other than self-selected non-participants, but is not likely to work well. In such a case, it would be essential that some elements of the characteristics that determine conditional participation probability not be included among the other explanatory variables included in the outcome equation. Otherwise, the conditional participation probabilities will be highly (multi-) collinear with these variables, resulting in a very imprecise estimate of the program impact. Strong candidates for variables to include in the participation equation, but not in the outcome equation -- variables that have a strong effect on the probability of participation but only a negligible direct effect on the outcome variable -- are hard to find.

We will return to this issue in the discussion of the methodology for a randomized outreach evaluation, where it is more critical.

A second feature of this methodology is the implicit assumption that program participation has the same impact for all participating fathers. This seems unlikely. A much more general model would specify entirely different relationships between the outcome variable and father characteristics for participants and non-participants; that is, participation would be modeled as changing the entire relationship between the baseline characteristics and the outcome variable, rather than a "parallel shift" of the equation.(7) Under this model the impact of program participation would vary with baseline characteristics in a very nonrestrictive way.

The sample sizes that would be required to obtain reasonably precise estimates of such a general model are not likely to be achieved given the size of current responsible fatherhood programs. We recommend, instead, that the assessment of variation in impacts with baseline characteristics be limited to examining interactions between impacts and a very small number of key characteristics, assuming that the effects of other baseline characteristics on outcomes are invariant to participation. This can be done by including as explanatory variables in the regression equation variables that are products of the conditional participation probability and selected characteristics of fathers, as discussed further in Appendix E.

B. Application to a Non-Experimental Design

In the non-experimental design presented in Chapter Three there is a group of volunteers from the target population for the responsible fatherhood program being evaluated -- the treatment group -- who may or may not choose to participate in the program and comparison population of fathers -- the comparison group -- who do not have the option of participating in the program. Thus, volunteers are in the treatment or comparison group because they are drawn from two separate populations; in contrast, in the experimental design study volunteers come from the same population and are randomly assigned to one group or the other. The absence of random assignment means that the characteristics of treatment group fathers likely differ from those of comparison group fathers in their baseline characteristics. The difference in mean outcomes for the treatment and comparison group fathers would presumably reflect differences in baseline characteristics as well as the program participation of some treatment group fathers.

The regression methodology described for the experimental case can be used to solve, or at least reduce, the problem caused by non-random assignment. The application of that methodology, including the use of estimated participation probabilities in outcome regressions, would be just the same is in the experimental design. In the non-experimental design, however, the other explanatory variables in the model serve to control for differences in baseline characteristics of treatment and comparison fathers, as well as to reduce standard errors. Hence, it is especially important to measure baseline characteristics that are important determinants of the outcome under this design. Confidence that the estimated program effect reflects the impact of the program, and not systematic differences in the baseline characteristics of the two groups of study volunteers, will depend on how well the evaluators can perform this task.

In many situations, the best statistical predictors of human behavior in a given period are proven measures of the same or similar behavior in previous periods -- employment this year is a much better predictor of employment next year than such variables as education, age, race, ethnicity, and family characteristics, for example. Hence, the outcome variable *measured at baseline* ought to be high on the priority list for explanatory variables to include in the outcome regression model. Thus, if the outcome variable is the child's score on a psychological test, there would be substantial benefit in testing the child at baseline as well as at follow-up.

C. Modifications for a Randomized Outreach Design

In the randomized outreach design (Chapter Three), study volunteers are randomly assigned to receive strong (treatment) or weak (control) outreach. Fathers in either group may decide to participate in the program, but the differences in outreach are expected to result in higher participation rates among fathers who receive the treatment outreach.

The regression methodology described for the experimental design can be applied to this type of design after making two modifications. First, as discussed in Chapter Seven, the conditional participation probabilities will be estimated from data for both the treatment and control subjects, one or more variables representing the randomized outreach will be key determinants of those probabilities. Second, the definition of the "modified participation variable" needs to be changed for the control group fathers. Recall that in the experimental design this variable is equal to the estimated conditional participation probability from the participation analysis for all treatment group fathers and zero for all control group fathers. In the randomized outreach design, the variable is the conditional participation probability for all fathers. These probabilities will be presumably be lower for control fathers than for treatment fathers, but they will not be zero, as in the experimental case.

In all other respects the model for the experimental design applies. With the modification in place, the estimated coefficient of the participation variable will be an unbiased estimate of the impact of the program on the outcome variable.

The role and importance of effective treatment outreach becomes evident by recognizing that this model is formally equivalent to a model discussed at the end of Section III.A, above, in which all volunteers are self-selected into participant or non-participant groups. We criticized that model on the grounds that the participation probabilities would likely be highly collinear with other explanatory variables in the outcome equation. The randomized outreach serves to break up this collinearity; the outreach variable would presumably be a key determinant of the participation probability, but would not be included among the other explanatory variables in the outcome equation.

The role of randomized outreach in the estimation methodology implies that the outreach must satisfy two important criteria. First, it must be effective; if it does not have a substantial impact on the probability of participation it will do little to reduce the collinearity between participation probabilities and other explanatory variables in the outcome equation. Second, it should have a negligible direct effect on outcomes. Some outreach methods might have substantial direct effects: efforts by respected role models to persuade fathers to participate and promises of long-term financial or other material rewards for participating are examples. Such methods might also be very effective in increasing participation, so some care must be exercised to avoid using them if the objective of random outreach is to help the evaluator separate impact effects from selection effects.

D. Extension to Categorical and Limited Dependent Variables

To this point we have assumed in our model specifications that the outcome (dependent) variable is a continuous variable with unlimited range. It is likely, however, that many key outcome variables will not satisfy both of these conditions. Some will be categorical (e.g. have the father and mother married) while others will have a limited range (e.g., hours of child contact and level of child support cannot be negative). Further, among categorical variables there are likely to be two types: qualitative variables, that indicate which of two unranked categories a father is in, and ordinal variables, where the categories have a meaningful ranking from lowest to highest (e.g., responses to questions that require selection of a value on, say, a five-point scale).

Appropriate modifications to the regression model can be made to accommodate each of these types of dependent variables. Possibilities include:

- Probit and logit for binomial dependent variables (qualitative or ordinal);
- Multinomial probit and logit for multinomial (more than two categories) qualitative dependent variables;
- Ordered probit for ordinal multinomial variables; and
- Tobit and many other limited dependent variable models for dependent variables with a limited range.(8)

The selection issue that is addressed in the context of regression analysis for the continuous unrestricted outcome variable assumed previously must also be addressed in these models. The approach to solving the problem is essentially the same as in the regression case. The evaluator could include an estimated participation probability variable, estimated from the participation analysis, as an explanatory variable in any one of these multivariate models. As in the regression case, however, the preferred estimation method is likely to involve joint estimation of the outcome and participation equations, by maximum likelihood or perhaps by some method that is less computationally intensive.

III. Extension to a Multi-Site Impact Analysis

In this section we begin by modifying the methodology discussed in Section II.A for the estimation of impacts in an experimental design for the evaluation of one program to the joint evaluation of multiple programs (including multiple sites for a single program). We assume that volunteers at each site are randomly assigned to control and treatment groups, that some treatment subjects do not participate in the program at each site, and that all control subjects do not participate. We also assume there is no cross-site contamination (e.g., subjects at one site participating in the program at another site.) We then turn to using the modified model in non-experimental and randomized outreach designs.

A. Experimental Design

Assuming for the moment that all treatment group fathers at all sites participate in the program, only two modifications to the regression methodology for the experimental single-site evaluation are needed. First, a set of "site dummies," variables distinguishing each site, should be added as explanatory variables in the regression. These will control for differences in the demographic, economic, and policy environments across the sites that are not captured by baseline characteristics of fathers. If the number of sites is very large, these could be replaced by a smaller set of variables that describe the environmental factors. While this would allow the evaluator to assess the effects of specific environmental factors on outcomes, it is unrealistic to expect meaningful results from such an analysis unless the number of sites is very large and the key environmental differences can be captured in a small number of variables.

Second, instead of using a single dummy variable to indicate whether the father is in the treatment or control group, the evaluator will likely use a separate dummy variable for each site. The coefficient of the treatment dummy for each site will be the estimate of the impact for treatment group fathers at that site. Estimates are likely to vary across sites because of variation in the way the programs are implemented, as well as for other reasons.

For any pair of sites, the difference in impacts can be estimated as the difference between the corresponding treatment dummy coefficients and a statistical for the null hypothesis of "no difference" can be easily performed. If the difference is not statistically significant, the evaluator may improve the precision of the estimates by constraining the estimated impacts for the pair of sites to be the same. This would be especially appealing for programs that are similar with respect to key program characteristics (e.g., multiple sites of a single program).

Of course not all treatment group fathers will participate, and the estimation procedure needs to be modified to take this into account. Analogous to the single-site case, the evaluator will need to replace the treatment dummy for each site by a modified participation dummy. For each site this dummy will have a value of zero for all control fathers at the site as well as for all fathers at other sites. For treatment group fathers at the site, the value of one should be replace with an estimated participation probability, obtained from the participation analysis (see Chapter Seven).

In Chapter Three we indicated that evaluating multiple sites would be one way to address the problem of small samples likely to be encountered in a single site evaluation. The gains are greatest if the programs' impacts and effects of other variables on outcomes are the same at all sites. Then, adding new sites is equivalent to increasing the sample at the first site. If the sites are sufficiently disparate in their programs, target populations, and environments, then there is no gain over conducting separate, single-site evaluations. The reality of any multi-site evaluation is likely to be somewhere in between. In selecting sites for a multi-site evaluation, homogeneous sites should be preferred over heterogeneous sites, other things equal, if improving estimator precision is a priority.

B. Non-Experimental Design

As in the single-site case, the methodology developed for the experimental design can be reasonably applied to the non-experimental design if careful attention is paid to measuring baseline characteristics that are predictive of outcomes. We assume that there would be a comparison group for each site and that each comparison group site would be matched to its corresponding treatment site on environmental characteristics that are likely to have an impact on outcomes. Under this condition, the site dummies in the model would capture the environmental factors common to each site.

An alternative would be to have a different, perhaps smaller, number of comparison sites than treatment sites. In the absence of matches for each sites, the site dummies would have to be dropped. They could be replaced with a set of variables that measure key aspects of the environment at each site, including the treatment sites (e.g., strength of the local labor market). The number of such variables would have to be small relative to the number of sites to obtain meaningful results.

C. Randomized Outreach Design

Under the randomized outreach design the specification for the outcome equation would be the same as under the experimental design except that the participation variable for each site would be set equal to participation probabilities for all fathers at the site, whether treatment or control, and to zero for fathers at all other sites. As discussed in Chapter Seven, the participation analysis itself would use data from both treatment and control fathers at all sites and the explanatory variables for that analysis would include both site and treatment dummies.

Return to ToC

---

1. It is also possible that attrition could be related to whether the subject is in the treatment or control group, regardless of participation, but this seems less likely.

2. The evaluator may want to estimate an attrition model analogous to that of the participation model for use in correcting possible attrition bias in the impact analysis. See Maddala, G.S. (1990) Limited Dependent and Qualitative Variables in Econometrics, Chapter 9, Cambridge: Cambridge University Press.

3. We assume the control fathers are not allowed to participate.

4. Referred fathers who do not participate may use some program resources in the recruiting process. While these may be small for each such father in comparison to resources used by participants, if there is a large number of such fathers, expenses incurred for their unsuccessful recruitment should not be neglected.

5. The appropriateness of this correction can be demonstrated mathematically as follows. Let p be the share of referred fathers who participate, let d be the mean effect of their participation (the quantity we are trying to estimate), let o_p be what their mean outcome would be if they did not participate, let o_n be the mean outcome for referred fathers who are non-participants, let o_t be the mean outcome for all treatment group fathers combined, and let o_c be the mean outcome for the control group. In the absence of the program, we would expect the control group and treatment group mean outcomes to be about the same (i.e., they would be the same except for random chance differences, which will almost certainly be small if the sample is reasonably large). The mean outcome for the treatment group would, in the absence of the program, be equal to:

$$p \times o_p + (1-p) \times o_n,$$

and this would approximately equal o_c. Because of the program, however, the mean outcome for the treatment group is:

$$o_t = p \times (o_p + d) + (1-p) \times o_n = p \times d + p \times o_p + (1-p) \times o_n = p \times d + o_c.$$

The last equality is approximate, based upon the expected relationship between the means for the treatment and control groups in the absence of the program. Subtracting o_c from the left- and right-hand sides of this equation and dividing by p yields the estimate described in the text:

$$(o_t - o_c)/p = d.$$

6. If the other explanatory variables explain little variation in outcomes, the standard error from the regression estimate may actually be higher than that for the difference in mean estimate, essentially because we have "wasted" information in our sample by trying to estimate the effects of some unimportant variables on outcomes.

7. See Maddala, op cit.

8. See Maddala, op cit.

CHAPTER NINE

SUMMARY AND CONCLUSIONS

I. Introduction

Below, we summarize the findings of our evaluability assessment, focusing on the findings from the site visits to the five fatherhood programs. We first restate the purpose of this report in Section II. In Section III, we describe features of the fatherhood programs we visited. In Section IV, we summarize where these programs are in terms of their readiness for a formal impact evaluation. We conclude in Section V with a discussion of critical next steps for fatherhood programs to improve their viability and evaluability.

II. Purpose of this Report

The increased interest in programs that promote responsible fatherhood and the limited information currently available on the services provided and effectiveness of these programs has generated interest in the systematic evaluation of responsible fatherhood programs. For this reason, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services and the Ford Foundation funded The Lewin Group and Johns Hopkins University to conduct an evaluability assessment of responsible fatherhood programs.

Fatherhood programs and emphasis on male parenting are relatively recent phenomena in the social service sector. Many of the programs currently in place are either very new or, if established, have been experimenting with new interventions or changing the program focus over time to meet the interests and objectives of funders. It is generally the case that fatherhood programs have not adequately documented their performance. This may be because of limited resources, a lack of experience with methods of measuring performance, or simply because the focus of program staff has been on serving fathers rather than proving that methods are effective. While program staff may believe that their activities are helping fathers and resulting in positive impacts on society, others, particularly funders, may be skeptical of evidence of program effectiveness that is limited to anecdotes.

Evaluations of responsible fatherhood programs can serve two important functions:

- provide information to outside agencies and organizations regarding the objectives and the effectiveness of their interventions, which may be used to attract and justify funding from these outside sources; and
- provide information to program staff that may be used to modify program design to more efficiently and effectively serve the fathers who use their services.

Systematic evaluation of fatherhood program outcomes is crucial to both program design and funding. Conducting rigorous evaluations using standard scientific methods can assist program operators in effectively planning their programs to meet funding requirements, in improving their work with fathers, and in furthering the development of the field of fatherhood research and policy. The goal of this report is to provide the Department of Health and Human Services and other policymakers with an evaluation design that can be used to evaluate a variety of responsible fatherhood programs. In addition, this report is intended to provide direction to organizations that would support or conduct evaluations by illustrating what is involved in the evaluation process and what mechanisms must be in place before a formal impact evaluation may be undertaken.

In developing this report, we conducted site visits of five fatherhood programs nationwide. The site visits allowed us to assess the readiness of fatherhood programs for formal evaluation, identify obstacles to their evaluation, and develop evaluation design alternatives that could be employed if such programs were to be formally evaluated in the future. The five programs selected for site visits were among those that had applied for funding from the U.S. Department of Health and Human Services, and are believed by DHHS staff to be representative of the more developed fatherhood programs in country.

III. Features of Current Fatherhood Programs

Individual fatherhood programs vary substantially in both the specific outcomes they attempt to achieve and the activities they undertake to achieve them. Among the five programs we visited, we observed substantial variation in the numbers of fathers served, the recruiting methods used, the services fathers received, and program goals. A common theme, however, was the underlying philosophy that in order to be an effective and responsible father, men needed first to develop the capacity to take care of themselves. Below, we describe some of the features of the five fatherhood programs we visited: the characteristics of participants, program objectives, service models, and sources of funding.

A. Characteristics of Participants

The majority of fathers served by the programs we visited shared a set of common characteristics. They were most often young (age 18 to 25), low income, African American fathers with a high school education or less. Fathers were generally unmarried and unemployed and had one or more children, most often under the age of five. Most of the programs served fathers who resided in the immediate geographic vicinity (neighborhood) of the program. Two programs served fathers on a county-wide basis.

B. Program Objectives

The programs we visited varied in terms of the specific outcomes each program was designed to affect. For example, one program has a particular focus on reducing infant mortality and improving child health by increasing the involvement of the father in pre-natal and child health care. This is a very specific objective not shared by the other fatherhood programs we visited. Another program, through its arrangement with the county court system, has increasing the level and consistency of child support payments as one of its primary objectives. This is only a secondary objective of the other programs we visited. There were, however, a number of objectives the programs did have in common. These include:

- increase education and employment;
- reduce alcohol and drug use;
- improve parenting skills;
- increase father involvement with his child(ren);
- improve attitudes or feelings toward children; and
- improve social and family interactions.

The above objectives represent those that fatherhood *program managers* believed to be the most important objectives of their programs. Through our conversations with government agencies and private funders we gained a sense of the objectives that they, as funders, believed to be most important for fatherhood programs to try to achieve. From the *funder's* perspective, the most important objectives include:

- reduce unplanned child-bearing;
- reduce criminal involvement;
- increase paternity establishment;

- increase contact with child;
- increase formal or informal child support;
- increase employment and earnings;
- increase education or training;
- improve child behavior; and
- increase cooperation with mother concerning child-rearing.

C. Service Delivery

1. Recruiting

The programs we visited used a variety of means to recruit and enroll fathers. Four of the programs relied heavily on outreach activities conducted by program staff, advertisements, and word of mouth to attract fathers. Two of the programs recruited fathers through contacts with either mothers or children participating in the primary programs offered by their sponsoring agencies. One program relied heavily on referrals from the county social service system, and another received nearly all of its participants via mandatory referrals from the county court system.

2. Participation

Three of the programs we visited have open-entry/open-exit participation policies. Fathers can participate on either a regular or irregular basis. These programs are experiencing difficulty defining exactly who is an active participant in their program. This is because a number of men in their programs do not participate on a regular basis, periodically returning to the program after long intervals of non-participation.

Two of the programs we visited have more defined intervals of participation. One program has a very strictly defined six-week curriculum. Full attendance in all program activities during the six-week period is mandatory in order to continue to be a participant in the program. The other program required court-ordered fathers working less than 32 hours per week to participate in program activities until they were able to pay child support for three consecutive months. After that, participation was optional. In this program, fathers often returned periodically both on a voluntary and involuntary basis.

3. Services Offered

The programs we visited offer a range of specific services with some of the services being very similar across the programs. What differs greatly, however, is the emphasis each program has on particular services and the manner in which they are delivered. For example, all programs offer some form of a men's support group and instruction in parenting. The primary focus of one program, however, is intensive in-home counseling. None of the other programs offer this service. The main focus of another program is classroom-style instruction using a uniform curriculum that covers a range of topics including black history, parenting, and job search and employment skills. The other three programs conduct more traditional case management activities, offering men's support groups and parenting instruction as core services and providing internal or external referrals for services such as GED preparation, employment training, substance abuse treatment, and help with child support enforcement or other legal matters on an as-needed basis.

D. Funding Streams

None of the programs we visited had a single, established, long-term source of funding. For most of the programs, funding comes from a variety of local community sources which often change over time. Such sources include: the county court or social welfare system, local foundations, the state, and private donations. All programs receive some portion of their funding through the Department of Health and Human Services. One of the programs receives nearly all of its funding through the federal Healthy Start program, and one program receives a share of its funding from a large national foundation.

IV. Findings From An Evaluability Assessment of Selected Programs

There are several important traits that programs must develop before a rigorous impact evaluation may be conducted. These include:

- measurable outcomes;
- defined service components and their hypothesized relationship to outcomes;
- an established recruiting, enrollment, and participation process;
- understanding of the characteristics of the target population, program participants and program environment;
- ability to collect and maintain information; and
- adequate program size.

Below, we describe where the fatherhood programs we visited are in their development of each trait. In general, the programs we visited appear not to be ready for a formal impact evaluation. This is due primarily to three factors: the programs are very new and still at the stage of refining recruiting methods and program services; the programs lack automated systems for tracking and reporting on clients; and the number of fathers served by most of the programs is very small.

A. Measurable Outcomes

Most of the fatherhood programs we visited were able to articulate a set of measurable outcomes believed to be influenced by the program. Among the most common were increased education and employment, reduced alcohol and drug use, improved parenting skills, and increased father involvement with his child(ren). Programs also cited some more difficult-to-measure outcomes, for example, improved attitudes or feelings toward children and improved social and family interactions.

One program had some difficulty defining a set of measurable outcomes influenced by program participation, mostly because the focus of the program was on general attitude change rather than on achieving more easily measured objectives. The primary goal of this program is to reconnect fathers with their children, or, in their words, "to turn the hearts of fathers to their children, and the hearts of children to their fathers." The underlying philosophy and secondary goal of the program is attitude change. Staff at this program believe that reconnecting fathers to their children will lead to changes in attitude and behavior leading to paternity establishment, job placement, and improved relationships with their child and the child's mother. Staff were, however, hesitant to identify specific consequences that could be used in an evaluation of their program.

B. Defined Service Components and a Hypothesized Relationship to Outcomes

Of the programs we visited, all were able to define the services they offered and, with the exception of the one program described above, link those services to hypothesized impacts on a set of measurable outcomes. The specific services offered tend to change over time, however. All programs seemed to be in the process of adding new services or refining those already in place. This is probably because most of the programs we visited are only a few years old.

C. Established Recruiting, Enrollment, and Participation Process

Of the programs we visited, most have established recruiting and enrollment practices. Only one program is in the process of experimenting with new recruiting techniques, as it is having difficulty attracting participants. This program also has a rather lengthy pre-screening process that would be difficult to replicate in recruiting control group members if an evaluation were to be conducted. With respect to program participation, two of the programs we visited are having difficulty defining exactly who is an active participant in their program. This is because a number of men in their programs do not participate on a regular basis, periodically returning to the program after long intervals of non-participation.

D. Understanding of the Characteristics of the Target Population, Program Participants, and Program Environment

All of the programs we visited seemed to have a good understanding of the population they serve and the environment in which the program operates. Many of the program managers live in or near the neighborhoods in which they operate their programs. While all but one of the programs lack an MIS, most of the programs still produce descriptive statistics on important characteristics of their participants, such as age, race, education, marital status, employment, number of children, and paternity status. In addition, most of the program managers we met seemed to be very knowledgeable about and well-linked to other agencies in the community such as state and local health and welfare agencies, child support enforcement, the criminal justice system, and agencies providing specific services to persons with low income such as housing, employment services, legal services, medical care, and substance abuse treatment.

E. Ability to Collect and Maintain Information

Only one of the programs we visited has any kind of computerized tracking system, and its system was still being developed and modified at the time of our visit. Another program has an MIS, but it is being used only to track female clients enrolled in its primary program. No computerized tracking of male clients is currently conducted.

F. Adequate Program Size

Most of the programs we visited serve a very small number of individuals, so it would be difficult for an evaluator to obtain statistically significant results. Only one program serves a relatively large number of fathers. The caseload of this program at the time of our visit was about 500 fathers. The program receives from 50 to 60 new referrals each month. This program is by far the exception. Three of the programs we visited serve only about 50 new fathers each year.

V. Next Steps

There are a number of steps that fatherhood programs can take to improve their viability and evaluability. Given the findings of this study, we suggest the following steps as being the most critical:

- ***Develop a core definition of what constitutes a responsible fatherhood program.*** The programs currently operating have arisen and evolved to respond to the unique needs of the communities in which they operate. There is considerable diversity across programs in both the objectives they try to achieve and the methods they use to achieve them. Establishing a minimum set of common objectives that define the basic mission of responsible fatherhood programs would be useful in furthering the research, development, and acceptance of these programs on the part of funders and policymakers.
- ***Conduct process evaluations.*** Programs should be encouraged to conduct process evaluations in order

to define program objectives and activities, identify best practices, and provide information that may be used to understand the important similarities and differences across the various programs. The use of a common process evaluation format would greatly facilitate the comparison of similarities and differences across programs.

- ***Build basic MIS capacity.*** An MIS is necessary to document a client's participation in the program, the services he receives and does not receive, and important outcomes related to program participation. The ability to track a client's progress through the program, both in terms of the services he receives and changes in important outcomes, is not only necessary before an evaluation effort can be undertaken, but is also useful to program managers who may use the information to improve program effectiveness. Without adequate documentation of program operations and performance, it can be extremely difficult for programs to obtain significant and stable sources of funding.
- ***Stabilize and enhance funding.*** Most fatherhood programs currently face a difficult paradox: they do not serve a sufficient number of fathers to support a formal impact evaluation, and they are unable to obtain sufficient funding to increase the number of fathers they serve because they have not demonstrated their effectiveness through formal evaluation. Programs may not initially be able to increase their sizes to accommodate a formal evaluation, but they can take a variety of steps in the direction of refining their goals and services, and documenting their effectiveness by less formal means in order to improve their performance and convince policymakers and funders of their usefulness. Stability of funding is important as well as levels of funding because changes to program objectives and services induced by the need to attract funds make it all the more difficult to establish a viable and evaluable program.

Return to ToC

APPENDIX A

LIST OF EXPERTS INTERVIEWED

I. Telephone Interviews

Martha Erickson

Director, Children, Youth and Family Consortium

University of Minnesota St. Paul, MN

Angela Greene

Senior Research Analyst

Child Trends

Washington, DC

Jean Grossman

Vice President and Director, Research and Evaluation Group

Public/Private Ventures

Philadelphia, PA

Kirk Harris

Project Director Center on Fathers, Families, and Public Policy

Chicago, IL

Jeffrey Johnson

Consultant

Management Plus

Washington, DC

Joseph Jones

Director, Men's Services

Baltimore City Healthy Start

Baltimore, MD

Kristen Moore Executive Director

Child Trends

Washington, DC

Edward Pitt

Associate Director, The Fatherhood Project

Director, National Practitioners Network

Families and Work Institute

New York, NY

Neil Tift

Director

Fathers' Resource Center

Minneapolis, MN

II. Participants at Practitioners Meeting

Charles Ballard

President

National Institute for Responsible Fatherhood and Family Revitalization

Washington, DC

Jerry Hamilton

Manager of Disadvantaged Programs

Goodwill Industries

Racine, WI

Joseph Jones

Men's Services Coordinator

Baltimore City Health Start, Inc.

Baltimore, MD

Wallace McLaughlin

Director

Fathers Resource Program

Indianapolis, IN

Ed Pitt Project Director

National Practitioners Network for Fathers and Families

New York, NY

Benjamin Powell

Program Coordinator

Inwood House Young Fathers Program

Bronx, NY

Barbara Kelly-Sease

Executive Director

Union Industrial Home for Children

Trenton, NJ

Neil Tift

Director

Fathers Resource Center

Minneapolis, MN

Return to ToC

APPENDIX B

SITE VISIT SUMMARIES

The Cleveland Institute for Responsible Fatherhood and Family Revitalization(1)

I. PROGRAM OVERVIEW

A. Background

The Cleveland Institute for Responsible Fatherhood and Family Revitalization (IRFFR) began services for fathers in 1982. The program is currently located at the Hough Center on Cleveland's east side, but serves men from all of Cuyahoga County. The program is funded by a variety of sources, including the State of Ohio, Cuyahoga County, the City of Cleveland Healthy Start program, the Cleveland Foundation, and a number of private sources.

The IRFFR staff we interviewed include:

- Charles Ballard, President;
- Stacy Hall, former Managing Partner in Cleveland but currently involved in the program's national expansion activities;
- Joanne Palmer and Ralph Moore, Managing Partners; and
- James Foster, Cheryl Foster, Albert Speis; Dale Powell, and Kenneth Austin, Outreach Specialists.

In addition to the IRFFR staff, we interviewed staff from the Juvenile Justice System (referring agency), and the Cleveland Foundation (funder).

B. Overall Goals of the Program

The primary goal of IRFFR is to reconnect fathers with their children. The underlying philosophy and secondary goal of the program is attitude change. IRFFR staff believe that reconnecting fathers to their children will lead to changes in attitude and behavior leading to paternity establishment, job placement, and improved relationships with their child and the child's mother. Additionally, fathers are encouraged to be self-reliant and not depend on staff for assistance. The IRFFR philosophy embraces the view that a father has the inner capacity to solve his own problems and the role of the program staff is to assist him through a process of self-discovery.

C. Characteristics of Participants

The IRFFR target population is primarily low-income, never married, non-custodial, African-American fathers who are 18-25 years of age. These fathers often have low skill and education levels. Currently, the IRFFR program service area is all of Cuyahoga County, OH (which includes the city of Cleveland). However, in an effort to better manage caseloads, staff are planning to narrow the catchment area to within a two-mile radius of the program site.

D. Services Provided

The core service provided to fathers who qualify under one of the IRFFR funding sources is in-home counseling. IRFFR staff, called Outreach Specialists, provide ongoing support to the father and are available 24 hours a day, seven days week. Outreach Specialists are required to wear pagers and respond to a father's call within 15 minutes of being paged. A significant amount of the Outreach Specialist's time is spent coaching the father to accept self-responsibility in becoming a meaningful part of his children's lives. On average, the Outreach Specialist spends about 30 hours per month (1-4 hours per visit) in the home of the father for a period of six months to a year, depending on the case. During the initial home visit, the father or "protégé" develops an annual plan, outlining goals he would like to achieve in the coming year. As discussed previously, Outreach Specialists provide no specific guidance or direction to fathers. Instead, they utilize an alternative counseling technique called *"creative questioning"* which enables the father to develop personal goals and identify his own resources (personal, family, community, etc.) to achieve them. Fathers are encouraged to be self-reliant and not depend on staff for assistance.

In addition to the intensive in-home counseling, all participants are required to complete a 16-week curriculum cycle. The participants attends group sessions based on the curriculum once per week. The topics discussed during the sessions include self-esteem building, fathering skills, health and nutrition, and male/female relationship building. All interested fathers can participate in the group sessions, whether or not they are formal protégés of the program. Not all interested fathers are able to find a funding source in order to obtain the in-home counseling services of the Outreach Specialist. In order for these individuals to receive the in-home counseling component, the program has established a modest self-pay arrangement, with rates based on the level of the father's income. Relatively few protégés, however, are funded in this manner.

IRFFR places a strong emphasis on staff role modeling. Staff are required to model a "risk free lifestyle" (no drugs or alcohol) as a condition of employment. IRFFR also seeks to hire married couples to serve as Managing Partners (program administrators) and Outreach Specialists. It is believed that program participants will emulate the married couple's behavior once they see a successful marriage modeled.

E. Recruitment/Enrollment/Participation/Completion

Recruitment

The IRFFR staff recruit fathers through outreach efforts designed so that program staff meet young fathers where they frequently gather in the community, including recreational centers, basketball courts, and playgrounds. Additionally, fathers are recruited through presentations at schools and churches, weekly group sessions held at the program site, and contract referrals from juvenile court, the County Child and Family Services, and Healthy Start. Most participants are referred to the program from the County (50-55%) and from Healthy Start (30-35%). A large number of the fathers are self-referrals, hearing about the program through "word of mouth."

Enrollment and Participation

The IRFFR program is open entry/open exit. The participant decides on his own whether he wants to participate. To be formally admitted however, the father has to qualify under one of the available IRFFR funding sources (Healthy Start or Child and Family Services). Even individuals who refer themselves to the program must be linked to a funding source in order to become a formal participant in the program. Once eligibility for a funding source is determined, the participant is assigned an Outreach Specialist who schedules a home visit with the participant. At this point, the participant becomes a protégé, the name used for formal participants in the program. Formal intake is done during the initial home visit. IRFFR typically serves approximately 100 protégés (one-third of whom are women) at any given time, and approximately

300 protégés per year.

Completion

There is no specified completion date. A father may continue to receive the case management services for as long as funding is available. If funding is not available, he may still participate in the group sessions.

II. EVALUATION ISSUES

A. Most Important Outcomes

IRFFR places greatest emphasis on changing the attitude of the father. Staff believe that once this is achieved, positive behavioral outcomes will occur including increased father involvement with his children, paternity establishment, and improved interaction with the mother of his children. In addition, IRFFR staff indicate that once a father has completed the program, he will demonstrate the following:

- Seek higher education;
- Seek suitable housing;
- Dress properly;
- Exhibit decorum changes;
- Become more articulate and use profanity less frequently;
- Improve his feelings toward his children; and
- Treat significant others with respect.

B. Data Availability

The IRFFR program maintains a number of forms and written notes on protégés in case files. With regard to protégés, a short (one page) intake form is completed, usually during the initial in-home visit. This form captures some basic demographic data about the individual -- age, ethnicity, marital status, last grade completed, employment status, legal concerns, and several other items -- as well additional data about other family members (e.g., name, whether paternity has been established, relation, date of birth, and address/telephone number). Other forms focus primarily on establishing participant goals and action steps needed to achieve the goals, and monitoring progress toward the goals. These forms include mostly handwritten notes. The number of contacts and hours of counseling is maintained for each participant, on a daily and monthly basis. Finally, outreach specialists maintain narrative notes within case files that document discussions with each protégé. These case notes are revealing of both the wide variety of problems encountered by participants and the courses of action mapped out in response to addressing each problem.

The IRFFR program is currently planning a new data system that would enable each outreach specialist to maintain data and case notes on a laptop PC. At the time of our visit, design work had not yet begun on the new system and no target date had yet been established for implementation of the new system. The only automated information maintained by the program is a spreadsheet which includes the following data items: name, address, zip code, telephone number, highest grade completed, date of birth, sex, income, funding source, whether paternity has been established, expiration date for funding, and the name of the outreach specialist.

C. Potential Evaluation Obstacles

Difficult to Measure Outcomes: Because the program focuses on attitude change it may be difficult to measure program outcomes. IRFFR staff were hesitant to describe any concrete program outcomes other than attitudinal change. They believe, however, that the attitude change leads to other positive consequences, such as greater child involvement, child support, paternity establishment, and employment. While the direct effect of the program, attitude change, may be difficult to measure, many of the proposed consequences of attitude change can be measured.

Small Sample Size: At any given time, the IRFFR program serves approximately 100 protégés, a third of whom are women. Over a year, the program typically serves about 200 men. These numbers correspond to formal protégés, and not to the number of persons who participate in group sessions but who do not receive the individualized services of the outreach specialists. IRFFR is in the process of developing program sites in other cities across the country. Inclusion of these sites in an impact evaluation would enhance the sample size.

The Baltimore City Healthy Start Men's Services Program(2)

I. PROGRAM OVERVIEW

A. Background

The Men's Services Program operates as part of the Baltimore City Health Start demonstration. The Baltimore City Healthy Start program is one of 15 sites nationally, funded with a five-year federal grant to provide intensive outreach and case management to women in areas at high risk of poor birth outcomes. Baltimore City operates Healthy Start programs at two sites: East Baltimore and West Baltimore. Each site covers five or six census tracts selected to participate based on economic and pregnancy outcome risk factors. The neighborhoods selected represent the poorest areas of Baltimore. The program is now in its sixth year of operation, as federal funding has been continued beyond the initial five-year grant.

The Men's Services Program (MSP) began as a pilot project in 1993 after recognizing the need for services for the men associated with the women and children involved in Healthy Start. MSP is funded almost entirely with Healthy Start funds, but has received small grants from private foundations. There are ten staff (one coordinator and four advocates at each site) dedicated to the MSP.

We interviewed staff at the West Baltimore MSP. The MSP and Healthy Start staff we interviewed include:

- Joseph Jones, MSP Director
- Karl Paige, MSP Assistant Director and Case Manager; and
- Peter Schafer, Healthy Start Policy Analyst.

B. Overall Goals of the Program

Because the goal of the Healthy Start program is to reduce adverse birth outcomes through increased prenatal, post-partum, and pediatric care, the primary aim of the MSP is to develop male parenting skills with the goal of reducing infant mortality and improving the health of children under the age of three. While the focus of the program is fetal and child health, MSP does, however,

attempt to address a variety of needs that low income, primarily non-custodial, fathers have. The program's goal may be more broadly stated as facilitating manhood development and the acquisition of life skills that are essential to effective fatherhood.

C. Characteristics of Participants

MSP participants are African American fathers of children age three and under participating in the Healthy Start program. They are between the ages of 17 and 35, mostly unmarried and unemployed. Over 90 percent of the mothers of their children are AFDC recipients. About sixty percent have a good relationship with the mother. Most of the men who come to the program do so for assistance with obtaining employment and developing parenting skills. The primary issues men have concern employment, substance abuse, custody/visitation, and relationships with their children and childrens' mothers.

D. Services Provided

MSP offers case management, a men's support group and fatherhood curriculum, parenting skills classes, family planning, GED classes, and a small employment initiative where fathers obtain jobs with private contractors involved in a lead-abatement program. The fatherhood curriculum emphasizes African American history and culture. The program has linkages with other organizations in the community to negotiate child support enforcement or other legal issues and for substance abuse treatment. The program offers transportation to and from program activities.

After an initial intake and assessment, case managers follow-up on fathers on a monthly basis to monitor the progress of fathers in attaining specific objectives developed after initial assessment. Fathers attend a men's group two evenings per week for two hours at a time. One of the sessions is devoted to the fatherhood curriculum developed by the program, and the other session is devoted to participants' specific issues. A hot meal is served at each session. MSP case management and men's group activities are conducted at a satellite site separate from where Healthy Start women's services are provided.

E. Recruitment/Enrollment/Participation/Completion

Recruitment

MSP participants are recruited through mothers who participate in Healthy Start. Neighborhood Health Advocates go out into the neighborhoods that comprise the program's target service area and knock on doors to talk to women and identify potential Healthy Start participants. A woman may initially enroll in Healthy Start if she is pregnant or has a child less than six months old. She and the child may then continue to participate until the child is three years old.

Healthy Start women are asked to identify their significant other or father of the child for participation in the MSP. Only about 50 percent of the women contacted are willing or able to provide contact information for a male significant other. This information is referred to MSP staff who then undertake outreach activities to enroll men into the program. Case managers make in-home visits and describe the services offered to men. MSP receives about four or five new referrals each month.

Enrollment

Once men are recruited to participate, they set up an appointment to meet with a MSP case manager for an initial assessment. The initial assessment involves obtaining the father's demographic, health and health care utilization, smoking, drug and alcohol use, education, employment history and employment barriers, family, contraceptive use, parenting/child development knowledge, and child support information. The information obtained is used to develop a plan for the father, called One Man's Plan, which lays out specific goals and objectives and a plan for achieving them based on the father's needs. All men who enroll in the program are associated with a mother in Healthy Start. About 70 percent of those who are referred by women to the program actually enroll.

Participation

Men participate by attending the weekly men's group sessions, meeting with their case manager to discuss monthly progress, and by participating in the other services to which they have been referred (parenting classes, employment, GED classes, substance abuse treatment, etc.). Currently, there are 200 men enrolled in the program (100 at each site), however, only a subset of them (maybe 50 percent) are active participants.

Completion

There is no defined completion date or duration for participation in the program. The program is currently trying to develop phases of participation: initial participation and an alumni group to act as mentors and facilitators. There are some participants who have been in the program for over two years.

II. EVALUATION ISSUES

A. Most Important Outcomes

MSP staff cited a number of specific outcomes that the program tries to achieve:

- increase the father's participation in the child's life, including participation in well-baby and early childhood health care visits;
- increase the father's level of education and employment;
- reduce rates of incarceration and criminal activity;
- reduce the level of substance abuse;
- improve the father's attitudes toward and relationships with his children and partner.

B. Data Availability

The MSP collects extensive information on fathers on the initial assessment form (described above), however, the information is not maintained electronically. The Healthy Start program does have the MIS capability to maintain such data. It collects and reports extensive information on mothers and children as part of a formal evaluation of the national demonstration. The system could easily be adapted to collect information on fathers. Currently, no follow-up information is collected on fathers, however, the program is considering doing a study to examine the impact of the program on employment and incarceration among participants. Such a study would require the collection of follow-up information on participants.

C. Potential Evaluation Obstacles

Small Sample Size: The primary obstacle to conducting an impact evaluation of the MSP is the program's small sample size. The two sites combined have served, to varying degrees, only 200 men over the last three years. However, if similar programs were adopted by other Healthy Start sites, sample sizes would likely become adequate.

Difficult to Identify a Comparison/Control Group: Currently, the program is not experiencing excess demand for its services. This, combined with the small number of participants precludes an experimental design. Finding a comparable comparison group may be difficult. For the Healthy Start evaluation, women residing in an adjacent geographic area were chosen as a comparison group, based on having similar risk factors to the program area. Since initiation of the project, however, the characteristics of the comparison group have changed (based on analysis of recent census data).

The Baltimore St. Bernadine's Male Involvement Project(3)

I. PROGRAM OVERVIEW

A. Background

The Male Involvement Project (MIP) is operated through the St. Bernadine's Head Start program, located in Baltimore, MD. MIP began in 1982 and currently receives funding from a number of sources including the US Department of Health and Human Services, the Rauch Foundation, the Families and Work Institute, the Annie E. Casey Foundation, and the Governor's Office on Alcohol and Substance Abuse. The program is located on the premises of St. Bernadine's church in central Baltimore.

The St. Bernadine's staff we interviewed include:

- James Worthy, Director of the Male Involvement Project;
- Sheila Tucker, Director of St. Bernadine's Head Start; and
- YaYa Robertson, Outreach Coordinator.

In addition to the St. Bernadine's staff, we interviewed two fathers who currently participate in the program.

B. Overall Goals of the Program

The primary goals of MIP are: (1) to link one male with each child participating in the Head Start program; and (2) to assist men in dealing with their needs so they may develop the capacity to care for children. The program does not necessarily focus on fathers. They hope to link a caring male role model (whether it be a father, friend, or other family member) with each child in their program.

C. Characteristics of Participants

Participants in the MIP are generally low-income, African American males between the ages of 19 and 35. Most have completed high school or have some skills training. About 70 percent have established paternity. Most participants (75%) have a child attending the St. Bernadine's Head Start program.

D. Services Provided

The primary services offered by the MIP include:

- Men's Support Group: MIP offers a men's support group that meets weekly for two hours. The group is directed by a contracted mental health specialist and by the program director. The purpose of the group is to discuss and resolve issues and problems important to the men. The group leaders bring a predetermined topic to discuss to the group each week, however, if there are other personal issues the men want to discuss that week, those issues take precedence.
- Referral to a variety of social service organizations: The program has relationships with many local social service agencies to which it refers participants for assistance with housing, food, substance abuse and mental health problems, medical care, and GED training.
- Economic Development Program: The program has begun to implement an employment training project whereby a local security agency provides an eight-week training course. Graduates from the

training course are hired by Head Start centers in Baltimore. In addition to the security training, the center also offers an Early Childhood Education certification course. To date, these programs have had very few participants, and even fewer graduates.

- Parent Training Curriculum: Beginning this year (1996/97), MIP will offer a structured curriculum for men that focuses on relationships and parenting. This program differs from the men's support group in that it will be more structured and informational. The classes will be held at least twice monthly with a different topic covered in each session.
- Periodic Special Events: The program also sponsors periodic dinners, father/child activities, and award ceremonies for participants.

E. Recruitment/Enrollment/Participation/Completion

Recruitment

Men are recruited through two primary means: (1) contact through mothers of children participating in the Head Start program, and (2) advertisement and word of mouth in the community. When children are enrolled in the Head Start program, the mother is asked to identify the father or other significant male that may be contacted to participate in the Male Involvement Program. The men are then mailed a brochure describing MIP and requesting a mail-in response if they are interested in participating. In addition, the MIP outreach specialist contacts these potential participants and attempts to enroll them in the program. Contact may be made through an in-home visit.

Men are also recruited through the advertisement of the Men's Support Group in the community. Flyers and word-of-mouth are the primary means for drawing individuals into the group.

Pre-Screening and Enrollment

No formal pre-screening of participants is conducted. Participants, along with the outreach specialist or program director, complete an intake form that provides basic information about the participant, including: name, address, phone, age, information on children, education, current employment status, marital status, guardianship status, whether or not there is communication with the mother, availability to participate in the program, and the services in which they are interested. For the current program year, a new client needs assessment form has been added that will solicit information on the participant's present family situation, substance abuse, employment history, barriers to employment, and personal goals.

Participation

Anyone who participates in the Men's Group is considered a participant. The Men's Group is the basic service through which individuals become involved in the program and may receive referrals to other types of services (listed above) and may be recruited to participate the Employment Development Project. To date, there have been very few participants in the Economic Development Project. Only two fathers have completed the security training component. Currently, about 7 to 10 fathers participate in the weekly Men's Support Group. Program staff indicated that they served (had one or more involvements with the program) about 125 individuals. Over the last three years, however, only about 50 men have been regular participants (primarily in the Men's Support Group).

Participants may also receive an in-home case management visit from the outreach specialist or program director. This is usually only done initially to enroll men in the program, but may also be done periodically if it is believed that the participant needs occasional follow-up.

Completion

There is no defined program completion date. With the exception of the employment programs, men may continue to participate in all activities for as long as they want. The employment training programs are for a defined duration (eight weeks for the security training and 90 hours for Early Childhood Education certification).

II. EVALUATION ISSUES

A. Most Important Outcomes

MIP staff indicated that the most important outcomes the program works to achieve are:

- reduce the consumption of drugs and alcohol;
- increase the education level (GED completion);
- increase the ability to find and maintain employment; and
- improve social and family interactions.

Staff indicated that, to date, paternity and child support have not been issues specifically addressed by the program.

B. Data Availability

MIP collects a variety of information on the initial intake forms including: age, education, place of residence, marital status, number of children and their ages, guardianship status, and whether or not there is communication with the mother; and current employment status.

Beginning this year, the program will collect additional information on the participant's present family situation, substance abuse, employment history, barriers to employment, and personal goals. MIP also keeps rosters of who participates in all activities sponsored by the program. All information is maintained in paper files.

C. Potential Evaluation Obstacles

Small sample: Probably the greatest obstacle to conducting a formal evaluation of MIP is the small number of individuals served by the program. MIP staff indicated that last year, 125 men had some involvement with the program. Very few, however, had ongoing involvement. Only about 50 men over the past three years have been regularly involved in the Men's Support Group, the primary service offered by the program. Only two men have completed the security training component of the program.

Defining service components and open-ended completion: Men may receive a variety of services through the program, either directly or through referrals. The Men's Support Group is the primary service offered to fathers. Most services, however, seem to be geared toward the Head Start mothers rather than specifically to fathers. Men may participate at any time for any length, with some returning only periodically. This lack of a defined service delivery would complicate efforts to define who is actually receiving program services.

Limited and informal tracking of participants: Currently, attendance rosters are kept for the weekly Men's Support Group. Files on fathers are not currently kept separately. Because of program funding, the information on fathers is included in an associated child's file in the Head Start program.

The Indianapolis Father Resource Program(4)

I. PROGRAM OVERVIEW

A. Background

The Father Resource Program (FRP) is operated through Wishard Memorial Hospital, located in Indianapolis, IN. The program has been operating since November, 1993 and currently receives funding from a number of sources including: The Lily Endowment (40%), Wishard Memorial Hospital (25%), The Indianapolis Foundation (15%), US Dept. of Health and Human Services (15%), Wishard Memorial Foundation (5%), and block grant funding from the Governor's office, Division of Children and Family Services (5%). The annual budget for the program is approximately $300,000. The program is located on the premises of Wishard Memorial Hospital in downtown Indianapolis.

The FRP staff we interviewed include:

- Wallace McLaughlin, Director;
- Frank Snyder, Project Social Worker;
- Kabir Sharif, Outreach Coordinator; and
- Carol Barber, Employment Developer.

In addition to the FRP staff, we interviewed Candace Curry from the Prosecutor's Office, and attorney Paul Malone who are also involved with the FRP program.

B. Overall Goals of the Program

The primary goals of FRP are: (1) to develop the capacity of young fathers to become responsible and involved parents, wage-earners, and providers of child support; and (2) assist fathers with developing the skills and behaviors necessary to cooperate in the care of their children, regardless of the character of the relationship with the mother. The overall goals may be restated as follows:

- Reducing welfare dependency, criminal involvement, and drug usage/selling;
- Enhancing young fathers' abilities to fulfill their roles as nurturing parents and providers.

A primary goal of the program is to place fathers in jobs upon completion of the program's six-week curriculum.

C. Characteristics of Participants

Participants in the program must be between the ages of 18 and 25 and the father of a child 3 years old or less, or an expecting father. The program serves primarily African American males (98%). About 80 percent have no high school diploma, 90 percent are unemployed, and about 65 percent have not established paternity upon entering the program. Initially, the program served individuals residing in the 20 square mile Blackburn area if the mother used Wishard Memorial Hospital. Currently, the program serves individuals from all of Marion County.

D. Services Provided

The fathers participate in six weeks of classroom instruction and discussion, and job counseling readiness, and placement activities. There are generally 8-10 participants in a class. At the end of the six-week curriculum, it is expected that the father will be placed in a job.

The six-week curriculum has several components:

- Instruction/discussion of black history, the definitions of 'boy', 'man', and 'father' and the role of the father in the family, community, and society. This part of the curriculum incorporates required readings from works such as Malcom X. Up from Slavery, and Visions of Black Men. Participants are also required to write their own autobiographies;
- Weekly elections for class leadership positions including a Class Leader (spokesman for the class); a Ritual Leader (leads daily meditation); and a Sergeant-at Arms (enforces rule of class, levies and collects fines for poor conduct);
- Instruction in parenting skills and child development (four 1.5 hour sessions) provided by staff from Family Services;
- Discussion/information/guidance on legal matters, including pressing civil/criminal legal concerns of participants;
- Information/counseling on paternity given by staff from the Prosecutor's Office;
- Information on AIDS and sexual responsibility;
- Field trips to local businesses, the library, voter registration;
- Speakers/role models from the community;
- Other activities such as trips to the YMCA for recreation each Friday, award ceremonies/dinners, and family nights;
- Job readiness instruction including filling out applications, taped mock interviews, developing a resume/work history, appearance, and problem solving; and
- Job placement - historically, to jobs in the Wishard Hospital, now more jobs found outside of the hospital. Jobs typically pay $6 or more per hour and provide benefits after a period of probation.

Participants in the six-week curriculum are paid a stipend, ranging from $75 to $115 per week, which is based on their performance. Performance criteria include: attendance, punctuality, attitude/conduct, appearance, and academic performance on assignments and exams. Participants may be fined from $0.50 to $6.00 per incident, depending on the infraction and may be dismissed from the program due to poor performance. Participants may also receive performance bonuses for good attendance, doing well on exams, and for serving as a class officer. Performance and stipends are determined weekly.

Participants attend classes Monday through Friday from 8:00 am until 3:30 p.m. Those enrolled in GED coursework attend GED classes daily from 5:00 p.m. to 7:00 p.m. In addition, there are 6 to 8 hours of family night activities during the six week period.

At the end of the six weeks, students who have successfully graduated from the program receive at $100 graduation stipend and are placed in a job. If they retain their job for 3 months, they are paid a $50 bonus, for 6 months, a $100 bonus, and for 1 year, a $150 bonus (for the same job) or $100 (for a different job).

E. Recruitment/Enrollment/Participation/Completion

Recruitment

When the program began, fathers were recruited primarily through a social worker employed by the hospital who identified potential participants through patients in the maternity ward. The social worker is no longer at the hospital and since her leaving, the program has not received many referrals from hospital staff. In general, nurses and other hospital staff do not seem interested in referring fathers to the program, even though program staff have made attempts to inform them about their services.

Currently, recruiting efforts have focused on the community at large and have involved radio ads; flyers posted in laundromats, car washes, restaurants, and churches; posters with a mail-in contact card placed in pool rooms, clinics, and other public places; and referrals from former program participants. Former students receive a recruitment bonus of $15 for referring someone who subsequently enrolls in the program. They receive an additional $25 if that individual completes the program or an additional $50 if he successfully completes the program (distinction discussed below).

Pre-Screening and Enrollment

Individuals interested in participating must undergo a rather intensive pre-screening application process. The pre-screen involves several interviews with FRP staff to inform the applicant about what the program involves, to determine how serious the applicant is about participating, and to assess the applicant's ability to participate and potential for successful completion of the program curriculum.

Interested applicants are interviewed on at least three occasions. During the first or second interview, the applicant's reading and writing skills are assessed. After the third interview, the applicant completes a detailed application form. A final interview is conducted with all program staff to clarify the applicant's criminal/drug history, the program intent, and other details on the application form.

About 50 individuals typically go through the pre-screening process. Of these, approximately half are accepted to the program. The FRP outreach coordinator cited reasons why applicants come to the program including: there are those who simply want the stipend, those who have nothing better to do, those who are just curious, those who think they want to change yet do not know what is involved, and those who truly desire to change and are willing to work for it. The arduous pre-screening process is intended to weed out those who are not truly serious about wanting to change. Less motivated applicants typically screen themselves out of the program once they learn exactly what it entails.

Following acceptance into the program, participants attend a week long orientation prior to officially beginning the program. The stipend begins during this week. During this orientation, students are advised of the performance criteria and given 40 to 50 pages of reading materials that will be discussed during the first official week of the program.

Participation

During the first week of the program, participants officially become a "student". The first week is considered a probationary period in which students may be dismissed if they do not pass a weekly exam covering the reading materials, or for poor attendance, poor attitude or generally poor performance. Students are also administered a drug test during the first week. After the first week, the class size has generally dwindled to approximately 8 to 10 students. As discussed above, the students participate in a six-week curriculum where weeks 1 through 4 focus on responsible fathering, and weeks 5 and 6 focus on job readiness and placement. Participants go through the program in a cohort and group dynamics play an important role. Students are encouraged to take leadership roles in their class by running for one of three class offices in weekly elections.

Completion

There are three levels of program completion:

- ***Successfully Complete***: Attend at least 80% of class sessions; complete all assignments on time;

minimal deviations from standards of class conduct, dress, and appearance; participate in post class activities (Recognition Ceremony, Community Service Day, and others); and pass second drug screening and the pre-employment drug screening.
- ***Complete***: Attend 70-80% of class sessions; complete all assignments; generally adhere to standards of class conduct, dress, and appearance.
- ***Participated***: Attend fewer than 70% of class sessions; complete some assignments; minimally adhere to standards of class conduct, dress, and appearance.

Students who successfully complete the program are "guaranteed" a job upon graduation. In the past, positions were typically found at Wishard Hospital. Currently, however, the program has an employment developer on staff who is attempting to make contacts with local employers who would be willing to hire FRP graduates. This has resulted in more placements outside of Wishard Hospital.

For those who only complete the program, FRP staff work hard to find a job, but placement is not guaranteed. The final "participated" category is currently being phased out. FRP now attempts to dismiss individuals from the program early on who are not willing to meet the standards for completion. Of the 8 to 10 who begin the program, about 6 to 8 graduate (complete or successfully complete).

II. EVALUATION ISSUES

A. Most Important Outcomes

FRP staff cited a number of specific measurable outcomes that the program works to achieve:

- Increase father's ability to provide financial support for himself and his child through job readiness and job placement activities;
- Increase father involvement with his child;
- Increase father involvement in the community;
- Increase father's education level (GED completion);
- Reduce father's involvement with drugs and other criminal activity;

- Paternity establishment;
-
- Help fathers to finish something they have started (i.e. the FRP curriculum).

B. Data Availability

FRP collects a variety of information on the initial application forms including:

- Demographic information: age, race, education, place of residence, living situation, marital status and other primary relationships, number of children and the ages, paternity/custody status, and AFDC participation status of each;
- Sources of income support, job training, skills, and interests;
- Criminal history, gun permit, and substance abuse information; and
- Expectations about what the applicant hopes to gain from the program.

In addition, weekly performance review and stipend information is also reported. The program maintains some of the demographic information and information on many of the outcomes described in the previous section on an electronic database. At this time, information is only maintained for students who actually enroll in the program. Plans are in progress to maintain information on individuals who apply but do not formally enroll in the program. FRP periodically summarizes information on program participation, attrition, and specific outcomes.

FRP is currently developing a follow-up database that will track outcomes for participants in the areas of paternity establishment, child support, arrears, visitation, employment, job duration, wages, educational attainment, and criminal activity. Follow-up information will be collected on former participants every six months.

C. Potential Evaluation Obstacles

No Excess Demand: FRP does not have a regular source of referrals and has not experienced an excess demand for their services. They have been able to accommodate all individuals who enroll and probably have the capacity to serve a greater number of individuals than are currently served. As discussed above, they have many individuals who initially are interested in the program, however much fewer actually enroll after learning what the program entails.

In the past, FRP staff have focused their efforts on curriculum development, currently they are focusing on ways to recruit participants. The lack of excess demand may create problems if a random assignment evaluation approach is adopted, however, their current recruitment strategies and experimentation with alternative strategies offers promise for a quasi-experimental evaluation approach.

Large, Ill-Defined Service Area: The program currently draws participants from a very large target area. The use of radio ads as a major source of outreach has increased the target area of the program substantially. This poses a problem if a non-experimental evaluation approach is adopted as it may be very difficult to identify a control group.

Small Sample Size: Class sizes are quite small (8 to 10) and the number of students who complete the program is even smaller (6 to 8 per six-week interval). Since the program began two and a half years ago, it has only served about 85 individuals. This implies that a relatively long period of observation may be necessary in order to obtain a sufficient sample size to conduct an evaluation.

Follow-up: Currently, the program has been able to maintain follow-up contact with about 25 to 50 percent of those who completed the program. In addition, the presence of the job retention bonuses may create a selection bias in the follow-up process, i.e. those with positive outcomes (those who retain jobs and seek the bonus) are more likely to maintain contact with the program.

The Racine Goodwill Industries Fatherhood Programs(5)

I. PROGRAM OVERVIEW

A. Background

Goodwill Industries of Southeastern Wisconsin operates several programs serving fathers in Racine County that fall under an umbrella program entitled "Children Upfront". The two primary programs serving fathers under Children Upfront include (1) Children First and (2) the Young Fathers Program. Children First, begun in 1990, is a program conducted in cooperation with the Racine County Human Services Department and the county Child Support Enforcement Agency. The Young Fathers program, begun in 1991, was part of the six-city Public/Private Ventures nationwide demonstration project. The Young Fathers program has become part of Wisconsin's Children First program. The Goodwill programs receive most of their funding from the State of Wisconsin and Racine County, and a small amount from private donations. The programs are located in downtown Racine, WI.

The Goodwill Industries staff we interviewed during the site visit include:

- Jerry Hamilton, Manager - Disadvantaged Programs;
- Craig Oliver, Case Manager for the Young Fathers program; and
- Michael McFarland, Case Manager for the Children First program.

In addition to the Goodwill Industries staff, we interviewed Christopher Lindroth, an attorney with the Racine County Child Support Enforcement Agency responsible for making referrals to the Goodwill programs.

B. Overall Goals of the Program

The main goals of the Goodwill fatherhood program are (1) to identify fathers who are non-custodial and reconnect them to their children; and (2) to facilitate the father's ability to be a provider for his children. A primary goal of the program is to engage the father in job seeking and employment activities and to ultimately increase the level and consistency of child support provided by the father.

C. Characteristics of Participants

Between 1990 and 1995, the Goodwill programs served about 2600 fathers. The average age of participants is 29 years, with 40 percent being between the ages of 16 and 25. Two-thirds of all participants have never been married. The average level of education is 11 years. The program serves primarily African American (58 percent) and white (26 percent) fathers. Ten percent of those who participate have been recently released from a correctional facility.

Relative to all participants, the 40 percent of fathers age 25 and under are more likely to be African American (68 percent), to have been released from corrections (15 percent), and to have never been married (80 percent).

The program serves fathers residing primarily in the five census tracts that comprise the inner city of Racine.

D. Services Provided

The Goodwill programs offer a variety of services, depending on the father's needs. There is not a structured curriculum for all fathers. A case manager identifies the needs of the particular father and develops a plan with him. The plan may include parenting and father responsibility courses, job readiness training, job search assistance, GED courses, referral to drug or alcohol treatment, and support group meetings.

E. Recruitment/Enrollment/Participation/Completion

Recruitment

The Goodwill programs conduct few outreach and recruiting activities in the community. Most program participants (85 percent) are referred to the program via the court system. Individuals referred by the courts are those in violation of child support agreements or are referred as a result of paternity suits brought by the county on behalf of a welfare mother. These individuals are required to report to the program within 48 hours of receiving the court order. The program does receive some referrals from the health department, schools, and other community-based programs. The program is just beginning to receive referrals from other participants in the program. Participants in the Young Fathers program are more likely to come from these sources as only about 50 percent are court ordered to the program.

Pre-Screening and Enrollment

There is little or no pre-screening of potential participants done either by the court system in referring persons to the program, or by Goodwill staff in accepting persons into the program. The court system views the program as a means to document "reasonable effort to find employment" in building cases against child support offenders. It is a jail alternative for child support enforcement purposes. As such, it is used for as many persons as possible. The courts refer from 55 to 60 fathers to the program each month.

Participation

Court ordered participants must schedule an initial appointment with a program case manager within 48 hours of court order. The father is sent an intake form to complete before his appointment. The form is intended to identify the fathers needs from the program, which are discussed at the first scheduled appointment. If a father fails to show up for his initial appointment, his failure to comply can result in incarceration.

When the father meets with the case manager, the case manager works up an employment development plan (EDP). Court ordered participants are required to engage in 32 hours of program/work/education/job search activities each week. Fathers working fewer than 32 hours per week are required to continue participation in the programs scheduled on his EDP.

Completion

Court ordered fathers are no longer required to participate in the program once they are able to pay child support for three consecutive months. Fathers may continue to participate if they choose. It is also common for fathers to be referred back to the program periodically on both a voluntary and involuntary basis. There is no set completion date for participants. Only a rather small proportion (about 30 percent) actually complete the 26-week parenting curriculum offered by the program. Employment and incarceration are the main reasons why participants do not complete the curriculum. According to program staff, about 30 percent of participants become employed right away (within 3 to 4 weeks) and therefore must report to the program for only the requisite three consecutive months; about 50 percent need some extra help and therefore participate somewhat longer than three months; and about 20 percent are hard core cases that never end up paying child support.

II. EVALUATION ISSUES

A. Most Important Outcomes

Goodwill staff cited a number of specific outcomes that the program tries to achieve:

- reduce the father's likelihood of child-bearing out of wedlock;
- improve the father's ability to obtain and maintain consistent employment;
- improve the father's ability to pay child support on a consistent basis;
- increase the father's involvement in the child's life; and
- improve the father's interactions with the mother(s) of his child(ren).

B. Data Availability

Goodwill industries collects a variety of information on the initial intake forms including:

- Demographic information: age, race, education, place of residence, living situation, marital status and other primary relationships, number of children and their ages, and paternity/custody status;
- Sources of income support, job training, skills, and interests;
- Criminal history, and substance abuse information; and
- Transportation and drivers license information.

The program also tracks the fathers' work, job search, and training activities while in the program. All information is maintained in paper files.

C. Potential Evaluation Obstacles

Program Used as a Deterrent: Participation in the program, at least for most participants, is not voluntary. The program itself is used as a "punishment" for not paying child support. As such, it may be more difficult to evaluate in the manner used for traditional interventions. Many participants choose to leave the program very quickly by obtaining employment and paying child support. Because of this, the "most successful" participants will be those who get the least services from the program. The program could be evaluated relative to a situation where jail was the only alternative. A comparison community would have to be selected for the control group in this situation because in Racine, all eligible persons are required to participate in the program and the courts expressed an unwillingness to participate in a randomized experiment. Participants who are not court-ordered to the program do not represent a large enough group to make up the entire evaluation sample (they represent only 15 percent of all participants).

Open-Ended Completion and Non-Uniform Services: Participants in the program receive very different services and participate for different periods of time. The types of services received, to a large degree, center around employment. Fathers who work 32 hours or more do not need to participate in program-sponsored activities. In general, employment hours replace program-sponsored activities. As discussed above, those who are most successful receive the fewest services from the program. An evaluation could be conducted, however, that addresses the effectiveness of specific program components. For example, the father responsibility/parenting skills components as separate from the employment and job skills components (an evaluation of the employment services is currently being conducted by the state). Unless there is a careful design and cooperation from the court system in the way they make referrals, there is the potential for spillover effects -- persons receiving only the employment services may come in contact with the program or participants in the program.

No Excess Demand: As discussed above, the court system refers as many persons as they can to the program. There appears to be no excess demand for the program, making random assignment difficult. Program staff did indicate, however, that there is excess demand for their services outside of their usual referral source (the courts). Were they able to obtain funding to conduct outreach and to provide services to these additional participants, they would have no problem recruiting volunteers for a randomized experiment. The effects of the program estimated in such an experiment might, however, be very different from the effects on fathers who are under court order to participate.

Return to ToC

---

1. The site visit was conducted January 31 and February 1, 1996 by Burt Barnow of Johns Hopkins University, Jeff Johnson of Management Plus, Gina Livermore of The Lewin Group, and John Trutko of James Bell Associates.

2. The site visit was conducted October 29, 1996 by Burt Barnow of Johns Hopkins University, Jeff Johnson of Management Plus, and Gina Livermore of The Lewin Group.

3. The site visit was conducted October 4, 1996 by Jeff Johnson of Management Plus, Gina Livermore of The Lewin Group, and John Trutko of James Bell Associates.

4. The site visit was conducted June 17, 1996 by Gina Livermore of The Lewin Group and Jeff Johnson of Management Plus.

5. The site visit was conducted September 10, 1996 by David Stapleton and Gina Livermore of The Lewin Group.

APPENDIX C

PROCESS EVALUATION INTERVIEW GUIDES

Discussion Guide for Responsible Fatherhood Project Directors and Sponsoring Organization Administrators

A. PROGRAM CONTEXT

1. Describe the following characteristics of the program:

a. Geographic area served by the program (i.e., boundaries of the service area).

b. Characteristics of the population in the program area (e.g., race/ethnicity, poverty population, single parent families living in poverty, educational attainment/school drop-out rate, substance abuse, criminal activity, and other relevant population characteristics).

c. Size and relevant characteristics of the target population to be served by the program (including both adults and children).

d. Labor market conditions (e.g., structure of the job market, unemployment rate, wages, availability of entry level/low skill jobs).

e. Other relevant environmental conditions that may affect program design, operations, or effectiveness (e.g., availability of other programs/services in the locality).

2. How have these environmental factors affected the design of the program?

3. a. How did the locality (i.e., service area) initially respond to the program (e.g., supportive, antagonistic)? Why?

b. As the program has evolved, how has the locality responded to the program? If there has been change, why?

c. What local resources has the program been able to draw upon (e.g., volunteers, services provided through other organizations, churches, facilities, etc.)?

B. PROGRAM DESIGN AND SERVICES

1. From your perspective, what is the overall mission of your agency? What are the major goals/objectives of your program? [Note: Order these goals in terms of their priority.]

2. Have the agency's program goals/objectives changed since the inception of the program? If so, how and why?

3. a. Identify major program components and/or services available through the responsible fatherhood program for the participants, their families, and the community as a whole.

b. Describe each of these program components/services in detail (e.g., non-traditional one-to-one counseling, group counseling, family outreach, fathering skills, health and nutrition information, medical and housing referrals, and career guidance) [Note: Where possible, provide any program documentation or literature detailing the component or service.]

Component 1:

Description of services:

Program goals addressed by this component:

Service eligibility criteria:

Estimated Number of Participants/Family Members Receiving the Service (Annual, Unduplicated Count):

Participants:

Family Members:

Others:

Frequency, duration, and interval of service:

Staffing:

Effects/Outcomes of the Service on Participants:

[Note: Continue same format until all program components/services have been detailed]

4. Please describe important linkages that your program has with other programs/organizations to refer program participants for additional services (not provided by your organization directly). For each linkage, provide the following information [Note: provide information for three most important linkages.]:

a. Name of the organization and how long the linkage has been in existence.

b. Types of services provided through the linkage.

c. Number and types of participants referred annually.

5. Do program components/services change during various times of the year (e.g., the summer months)? If so, how?

6. Has the program encountered problems in retaining participants (and other family members) within specific services (e.g., workshops, counseling sessions, social and recreational activities, etc.) until they successfully complete the service? If so, why?

7. What is the level of family involvement in the program? What, if any, services are provided to families?

8. What aspects of the service delivery strategy are most innovative? Why?

9. Which components/services have been most/least helpful for program participants?

10. Were there services that you provided in the past that you have discontinued? If yes, which services and why were they discontinued?

11. a. Do you feel that the package of services offered (or are able to refer participants to) comprehensively meets the needs of your participants?

b. Is there excess demand for services? If so, for which services is there excess demand?

c. What service gaps exist and how could each of these gaps be addressed?

d. Are there other approaches, strategies, or services that would contribute to better outcomes for program participants?

C. PARTICIPANT RECRUITMENT AND PROGRAM INVOLVEMENT

1. a. How do individuals/families generally hear about your program and the services it offers (i.e., discuss specific outreach methods that are used by the program)?

b. Do you get direct referrals from other human service agencies? If so, which agencies and about what proportion come from each referring agency?

2. a. Have outreach/recruitment activities been targeted on particular types of individuals/families within the community?

b. Are there particular types of individuals/families who have been very difficult to reach and/or recruit?

3. To what extent has the program been successful in making individuals/families aware of the program within the community? What have been the keys to heightening awareness of the program?

4. Why do individuals/families come to the program (i.e., what are they looking for and what is it about the program that seems to attract them)?

5. a. Does the program use screening criteria to determine which individuals/families it will serve? If so, what criteria have been used?

b. What are the reasons that some individuals/families eligible for services do not subsequently become participants (e.g., have those not participating been screened out by the agency or selected themselves out)?

6. At what stage does the individual move from being considered a potential recruit to becoming enrolled as a participant?

7. a. Once participants are enrolled at the program, how long do they remain as participants (e.g., longest, shortest, and average duration in weeks or months)?

b. Has attrition from the program been much of a problem? If so, what are some of the reasons that participants terminate from the program?

D. CLIENT CHARACTERISTICS

1. Who are the individuals receiving services? To the extent possible, please provide a demographic profile, numbers served, and scope and intensity of presenting problems among the fathers served. [Note: The evaluator will also collect characteristics data through the data systems and records maintained by sites.]

2. Who are the fathers/families not receiving services? Were there specific groups of fathers/families within the community that the program did not serve? If so, why? What strategies might be employed to reach these individuals?

3. What factors (e.g. recruitment strategies, types of agencies coordinated with, particular types of services offered, local conditions) influence the types of individuals/families served and not served?

4. What are the characteristics of participants that drop out of the program after enrollment? When do participants usually drop out and for what reasons?

E. SERVICE INTEGRATION

1. What is the universe of organizations that provide or influence the delivery of services to the target population in the local community?

2. What historical linkages existed among service delivery organizations prior to the start-up of the program?

3. How has the responsible fatherhood program affected the degree of service integration/coordination in the delivery of services to the target population within the community? Describe any linkages with other community employment, training, education, health, public health, mental health, juvenile justice, and social services agencies/programs.

4. How have services been integrated/coordinated? Who is responsible for case management/coordination? What is the process of case management? What are the follow-up procedures for determining whether inter-agency referrals result in the provision of services? What are the outcomes for participants of these referrals?

5. What have been the advantages of coordination (e.g., reduced duplication of services, ability to provide a wider range of services, ability to better target services on the needs of clients, enhanced ability to recruit program participants)?

6. What have been the major barriers to coordination? Which services have been most difficult to coordinate?

7. Are there any clear service gaps that you have not been able to address through coordination?

F. PROJECT STAFFING AND STAFF DEVELOPMENT

1. For the current year, please provide the following:

Total Number of Full-Time Paid Staff at Site:

Total Number of Part-Time Paid Staff at Site:

Total Number of Full-Time Equivalent Staff at Site:

2. What is the staffing configuration of the program, including roles and educational levels? Identify each paid staff member and their role/function.

3. To what extent are volunteers used? How many volunteers are used? What role/function do they play (e.g., mentoring, counseling, administration)? How are volunteers identified and recruited?

4. Are staff development activities available? Describe the extent and content of staff/volunteer training provided.

G. PERCEPTIONS OF PROGRAM OUTCOMES/IMPACTS

1. What kind of overall effects has the program had on participants, their families, and the community as a whole? To date, has the responsible fatherhood program had specific effects in any of the following areas (Note: See impact study for specific measures under each type of effect) and, if so, how:

a. Responsible Father Behavior (e.g., safe sex behavior, reduced unplanned child-bearing, marriage/stable relationships, reduced substance abuse, reduced criminal involvement, and community connectedness).

b. Father's Relationship with Child (e.g., paternity status, contact/visitation, type of child-related activities in which the father participates, parenting skills, and closeness).

c. Father's Financial Capabilities/Support (e.g., child support, employment and earnings, work ethic/attitude, education/training activities, housing, other responsibilities, physical health, mental health, self-awareness/self-esteem, anger management, and ability to deal with racism).

d. Child Well-Being (e.g., safety in the household, physical health, emotional/mental health, academic achievement, social behavior, and problem behavior).

e. Co-Parenting Relationship (e.g., arrangement for child access, agreement on child support, agreement/cooperation concerning child-rearing, parents' feelings toward each other, father's attitudes toward significant others, and quantity and quality of communication between parents).

f. Other Perceived Impacts/Effects.

2. Has the program been responsive to the individual needs and desires of participants and their families? How has this been ensured?

3. Please provide any impressions that you might have about the program's impacts:

a. What aspects of your program's approach or services appear to contribute most to successful participant outcomes?

b. Are there particular types of individuals/families for which the program has been especially effective?

c. Are there particular types of individuals/families for which the program has been ineffective?

d. Are there characteristics of individuals entering the program that are likely to influence outcomes?

4. How have participant impacts/outcomes for the current program year compared to previous years' outcomes? What might explain any differences?

5. a. To what extent has the program been able to meet the needs of the surrounding community?

b. Can you identify any specific impacts that it has had on the surrounding community?

6. a. If a long-term study of your program was undertaken, do you have any suggestions about potential control or comparison groups (i.e., that would not receive the intervention, but from whom you might be able to gather data on characteristics and outcome measures)?

b. Do you think it would be possible to randomly assign individuals to treatment and non-treatment groups (i.e., withhold the treatment from individuals)? Why or why not? If yes, do you have any suggestions about how random assignment might occur?

7. a. What types of data are currently being collected on each participant? [Note: obtain a complete set of forms that are completed from the point of first contact to the last contact with the participant.]

b. What types and when does follow-up occur? For how long are participants tracked? For how long should participants be tracked?

c. Are data being entered into an automated participant tracking system? If so, get a listing of the variables.

H. PROGRAM FUNDING AND COSTS

1. If available, please provide a summary of the program's funding sources (for the most recent fiscal year _____):

Funding Source 1 Amount

Funding Source 2 Amount

Funding Source 3 Amount

Funding Source 4 Amount

Total Funding Amount:

2. What types and amounts of in-kind contributions are received by the program (e.g., donated

space, equipment, volunteers)?

3. What were the major costs involved in program start-up (by major category)?

4. What are the major ongoing costs for the program (e.g., staff, equipment purchase or rental, transportation, subcontracts, utilities, security, etc.)? [Note: collect a budget and expenditures to date.]

5. How do costs break down by major program component/service?

6. How do the types of participants served affect costs? What types of participants are most/least costly to serve?

7. Have certain services been more costly to provide than expected? If so, why?

8. What types of system(s) does the program use to track program expenditures/costs?

I. PROGRAM REPLICABILITY

1. What features of the responsible fatherhood program would be easiest to replicate in other localities across the country? What features would be hardest to replicate?

2. How do location, demographics, and other distinctive features at this site make the program either non-transferable or limit its transferability?

3. What needs to be communicated to other agencies involved in providing services for fathers (and families) in order for the responsible fatherhood program to be successfully transferred?

4. If you were to set up a new responsible fatherhood program in another community? How would it differ from what was done here?

Discussion Guide for Responsible Fatherhood Program Managers and Staff

A. PROGRAM CONTEXT

1. Describe the following characteristics of the program area:

a. Geographic area served by the program (i.e., boundaries of the service area).

b. Characteristics of the population in the program area (e.g., race/ethnicity, poverty population, single parent families living in poverty, educational attainment/school drop-out rate, substance abuse, criminal activity, and other relevant population characteristics).

c. Size and relevant characteristics of the target population to be served by the program (including both adults and children).

d. Labor market conditions (e.g., structure of the job market, unemployment rate, wages, availability of entry level/low skill jobs).

e. Other relevant environmental conditions that may affect program design, operations, or effectiveness (e.g., availability of other programs/services in the locality).

2. How have these environmental factors affected the design of the program?

3. a. How did the locality (i.e., service area) initially respond to the program (e.g., supportive, antagonistic)? Why?

b. As the program has evolved, how has the locality responded to the program? If there has been change, why?

c. What local resources has the program been able to draw upon (e.g., volunteers, services provided through other organizations, churches, facilities, etc.)?

B. PROGRAM IMPLEMENTATION

[Note: Ask staff member if he/she has been involved in the program since near the start-up of the program.]

1. Were you involved in the getting the program up and running? What was your role?

2. What factors facilitated project implementation? What barriers were encountered during implementation? How were barriers overcome?

3. What changes were made in the program during implementation and in response to what circumstances? Were any components or elements of the original program design not implemented or abandoned early on? Why?

C. PROGRAM COMPONENTS/SERVICES

1. a. Identify major program components and/or services available through the program for the participants, families, and the community as a whole.

b. Describe in detail those program components/services that you are regularly involved in (e.g., non-traditional one-to-one counseling, group counseling, family outreach, fathering skills, health and nutrition information, medical and housing referrals, and career guidance) [Note: Where possible, provide any program documentation or literature detailing the component or service.]

Component 1:

Description of services:

Program goals addressed by this component:

Service eligibility criteria:

Estimated Number of Participants/Family Members Receiving the Service (Annual, Unduplicated Count):

Participants:

Family Members:

Others:

Frequency, duration and interval of service:

Staffing:

Effects/Outcomes of the Service on Participants:

[Note: Continue same format until all program components/services have been detailed]

2. Please describe important linkages that your program has with other programs/organizations to refer program participants for additional services (not provided by your organization directly). For each linkage, provide the following information [Note: provide information for three most important linkages.]:

a. Name of the organization and how long the linkage has been in existence.

b. Types of services provided through the linkage.

c. Number and types of participants referred annually.

3. Do program components/services change during various times of the year (e.g., the summer months)? If so, how?

4. Has the program encountered problems in retaining participants (and other family members) within specific services (e.g., workshops, counseling sessions, social and recreational activities, etc.) until they successfully complete the service? If so, why?

5. What is the level of family involvement in the program? What, if any, services are provided to families?

6. What aspects of the service delivery strategy are most innovative? Why?

7. Which components/services have been most/least helpful for program participants?

8. Were there services that you provided in the past that you have discontinued? If yes, which services and why were they discontinued?

9. a. Do you feel that the package of services offered (or are able to refer participants to) comprehensively meets the needs of your participants?

b. Is there excess demand for services? If so, for which services is there excess demand?

c. What service gaps exist and how could each of these gaps be addressed?

d. Are there other approaches, strategies, or services that would contribute to better outcomes for program participants?

D. OUTREACH AND INTAKE

[Note: Ask these questions only if the individual is involved in intake.]

1. How is outreach to the target population conducted? [Note: Obtain available brochures or leaflets that may be used for outreach.] How are fathers identified and selected to participate in the program?

2. Do eligibility or admission criteria for services create barriers to access or enhance access? Are any types of incentives used to encourage participation (e.g., recreational activities, help with transportation, snacks, etc.)?

3. Are there cultural characteristics of the target group or the sponsoring organization that facilitate or create barriers to enrollment of the target group (e.g., language, ethnic background, or race)?

4. Why are some eligible fathers not receiving services? Which barriers to service are internal and which are external?

5. How are intake, participation, and/or enrollment defined?

6. What information is collected at intake? How much burden is placed on participants at the time of intake/enrollment and does this burden affect willingness to participate or the types of program participants?

7. Overall, how successful has the program been in recruiting participants? What factors account for success or failure?

E. ASSESSMENT AND CASE MANAGEMENT

[Note: Ask these questions only if the individual is involved in client assessment or case management.]

1. Once enrolled, how are the service needs of participants and their families determined? What is the process of matching service provision to participant/family needs? What are the most common service needs?

2. Who receives a service plan? How is the plan developed? Who is involved in the process (e.g., the participant, case managers, family members)? How is the plan updated and monitored?

3. How and when are participants assigned to case managers? If the interviewee is a case manager: What is his or her case management caseload? What is the typical caseload managed by a case managers?

4. Overall, how successful has the program been in tailoring services to the specific needs of program participants? Why?

F. CLIENT CHARACTERISTICS

1. Describe any distinctive characteristics of the fathers/families served by the program?

2. If the staff member is a case manager: To the extent possible, please provide a demographic profile, numbers served, and scope and intensity of presenting problems among the fathers and families you serve.

3. Were there specific groups of fathers/families that the program did not serve? If so, why? What strategies might be employed to reach these individuals?

4. What factors (e.g., recruitment strategies, types of agencies coordinated with, particular types of services offered, local conditions) influence the types of participants served and not served?

5. What are the characteristics of participants that drop out of the program after enrollment? When does dropout usually occur and for what reasons?

G. SERVICE INTEGRATION

1. Do you refer participating fathers and families to other agencies for services? Which agencies? Do you have a contact person at each agency? Is there a formal agreement to coordinate or are the arrangements informal? If formal, what are the key provisions included in the agreement?

2. How does the referral process work? Do you follow-up after you make the referral? How do you know the participant or family is getting the service?

3. Which have been the easiest agencies to work with? Which have been the most difficult?

4. Are there any clear service gaps that you have not been able to address through coordination?

H. STAFF DEVELOPMENT

1. What types of staff development activities have you been involved in? How helpful have these activities been?

2. Are there areas in which you feel that additional staff development is needed?

I. DATA COLLECTION AND RETRIEVAL SYSTEM

1. Do you currently complete any client forms? If yes, have these forms been satisfactory? Have they provided the types of information needed to effectively assess, case manage, and track program participants? How burdensome have these forms been for participants and staff? Has it been necessary to supplement these forms with other forms? If so, please describe these supplemental forms.

2. Has the automated client information system developed for this program been satisfactory? How burdensome has the system been for you? Has it been necessary to supplement this system with other data systems or files? If so, please describe these supplemental systems/files and provide copies of their data structures. Has the system provided automated reports that have been helpful in processing clients?

3. Has the automated system provided the types of output reports needed to effectively manage the program and/or case manager caseloads? Which of these reports are being used by you and how are you using them (e.g., tracking service delivery, assessing client needs, tracking outcomes)? How

could these reports be improved?

4. Do you have any suggestions on how the automated system might be improved to provide better tracking of participants and their outcomes?

5. What types of confidentiality problems are involved in maintaining data on program participants? Are there ways to resolve these problems?

6. Do you have any suggestions on how to best track participants longitudinally (i.e., through program involvement and beyond) and what types of data should be tracked? For what period of time do you believe it is possible to track clients?

J. PERCEPTIONS OF PROGRAM OUTCOMES/IMPACTS

1. What kind of overall effects has the program had on participants, their families, and the community as a whole? To date, has the responsible fatherhood program had specific effects in any of the following areas (Note: See impact study for specific measures under each type of effect) and, if so, how:

a. Responsible Father Behavior (e.g., safe sex behavior, reduced unplanned child-bearing, marriage/stable relationships, reduced substance abuse, reduced criminal involvement, and community connectedness).

b. Father's Relationship with Child (e.g., paternity status, contact/visitation, type of child-related activities in which the father participates, parenting skills, and closeness).

c. Father's Financial Capabilities/Support (e.g., child support, employment and earnings, work ethic/attitude, education/training activities, housing, other responsibilities, physical health, mental health, self-awareness/self-esteem, anger management, and ability to deal with racism).

d. Child Well-Being (e.g., safety in the household, physical health, emotional/mental health, academic achievement, social behavior, and problem behavior).

e. Co-Parenting Relationship (e.g., arrangement for child access, agreement on child support, agreement/cooperation concerning child-rearing, parents' feelings toward each other, father's attitudes toward significant others, and quantity and quality of communication between parents).

f. Other Perceived Impacts/Effects.

2. Has the program been responsive to the individual needs and desires of participants and their families? How has this been ensured?

3. Please provide any impressions that you might have about the programs impacts:

a. What aspects of your program's approach or services appear to contribute most to successful participant outcomes?

b. Are there particular types of individuals/families for which the program has been especially effective?

c. Are there particular types of individuals/families for which the program has been ineffective?

d. Are there characteristics of individuals entering the program that are likely to influence outcomes?

4. How have participant impacts/outcomes for the current program year compared to previous years' outcomes? What might explain any differences?

5. a. To what extent has the program been able to meet the needs of the surrounding community?

b. Can you identify any specific impacts that it has had on the surrounding community?

K. PROJECT COSTS

1. Please provide a breakdown of the percentage of time that you spend in a given month by type of activity.

2. How do the types of participants served affect costs? What types of participants are most/least costly to serve?

3. Have certain services been more costly to provide than expected? If so why?

L. PROGRAM REPLICABILITY

1. What features of the program are most and least replicable in other localities across the country?

2. How do location, demographics, and other distinctive features at this site make it either non-transferable or limit its transferability?

3. What needs to be communicated to other agencies involved in providing services for fathers (and families) for this program to be successfully transferred?

Discussion Guide for Community Human Service Providers

1. What are the major problems/issues faced by the local community? Have characteristics of the community or problems changed over the past five years?

2. Who are the fathers/families in need of services and how are they affected by conditions within the community? To the extent possible, provide a demographic profile, numbers in need, and the scope and intensity of presenting problems.

3. What characteristics of the community culture are important to understand in assessing the responsible fatherhood program and its results (e.g., neighborhood organization, ethnic groups, mobility of the population)?

4. What physical characteristics of the community affect problems within the community and availability of services (e.g., transportation systems, neighborhood boundaries)?

5. What other initiatives (other than the responsible fatherhood program) in the community serve the target population for the responsible fatherhood program (e.g., law enforcement initiatives, health care programs, welfare initiatives, grass roots movements)? What organizations are involved in these initiatives and what services are provided?

6. Do you know the goals/objectives of the responsible fatherhood program? If yes, what do you view as the major goals of this program? Do you feel that these goals are appropriate? Do you think the responsible fatherhood program is addressing these goals?

7. What are the goals of your agency/program? How do these goals relate to the goals of the responsible fatherhood program?

8. How and why did your agency come to be involved with the responsible fatherhood program? Did your agency face any barriers to working with the responsible fatherhood program? How were these barriers overcome?

9. Describe the current coordination arrangements. How are the arrangements maintained? Do you have a formal agreement with the responsible fatherhood program?

10. How many responsible fatherhood participants are being referred to your agency for services on a monthly basis? To the extent possible, please provide a demographic profile of the participants being referred for services.

11. How do responsible fatherhood program referrals compare to other types of individuals served by your agency?

12. Do individuals who are referred to your agency by the responsible fatherhood program come in for services? Do you have any follow-up procedures for determining whether referrals resulted in provision of services and what were the outcomes for the referred individuals? Have responsible fatherhood staff followed-up on the referrals they have made?

13. Please describe the specific services you provide for referred responsible fatherhood participants.

14. How successful are these referred clients in completing services? Are the patterns of service receipt and completion the same for other individuals who you serve?

15. How are clients helped by the services they receive through your agency? Can you point to any specific client outcomes? Do there appear to be any particular types of responsible fatherhood referrals that are being helped more or less by your services (in terms of specific outcomes)?

16. What impact has the collaboration with the responsible fatherhood initiative had on your agency?

17. Do you have any views about the relative impacts of the responsible fatherhood initiative in any of the following areas:

a. Responsible Father Behavior (e.g., safe sex behavior, reduced unplanned child-bearing, marriage/stable relationships, reduced substance abuse, reduced criminal involvement, and community connectedness).

b. Father's Relationship with Child (e.g., paternity status, contact/visitation, type of child-related activities in which the father participates, parenting skills, and closeness).

c. Father's Financial Capabilities/Support (e.g., child support, employment and earnings, work ethic/attitude, education/training activities, housing, other responsibilities, physical health, mental health, self-awareness/self-esteem, anger management, and ability to deal with racism).

d. Child Well-Being (e.g., safety in the household, physical health, emotional/mental health, academic achievement, social behavior, and problem behavior).

e. Co-Parenting Relationship (e.g., arrangement for child access, agreement on child support, agreement/cooperation concerning child-rearing, parents' feelings toward each other, father's attitudes toward significant others, and quantity and quality of communication between parents).

f. Other Perceived Impacts/Effects.

18. Do you have any suggestions about how the linkages between your program and the responsible fatherhood program might be improved?

Discussion Guide for Organizations Providing Funding and Oversight for the Responsible Fatherhood Program

1. Is this particular program <identify the specific program being evaluated> part of a larger initiative in the area of responsible fatherhood? If so, please describe this larger initiative and its goals.

2. Is this the only program site you are funding? If not, please list others. What is the total level of funding across all of your fatherhood initiatives? What is the level of funding that you provide for this particular responsible fatherhood program?

3. Why did your agency provide funding for this program? What were the features of the program that weighed most heavily on your decision to fund this program?

4. From your perspective, what are the major goals/objectives of the program? [Note: Order these goals in terms of their priority.]

5. How have these goals fit into your overall goals for your fatherhood initiative (if in fact this is not the only site)? How do the goals of this program compare to those of other fatherhood initiatives your organization has funded?

6. Are there any important contextual or environmental factors to take into consideration when assessing the outcomes of this responsible fatherhood program, such as:

a. Geographic area served by the program (i.e., boundaries of the service area).

b. Characteristics of the population in the program area (e.g., race/ethnicity, poverty population, single parent families living in poverty, educational attainment/school drop-out rate, substance abuse, criminal activity, and other relevant population characteristics).

c. Size and relevant characteristics of the target population to be served by the program (including both adults and children).

d. Labor market conditions (e.g., structure of the job market, unemployment rate, wages, availability of entry level/low skill jobs).

e. Other relevant environmental conditions that may affect program design, operations, or effectiveness (e.g., availability of other programs/services in the locality).

7. Please identify major program components and/or services that your agency contracted with the fatherhood program to provide to the target population? [If possible, please provide a copy of the contract or written agreement that sets forth the scope of work under the project.]

8. What is your assessment of how well the responsible fatherhood program has provided these contracted services? Are there program areas of particular strength or weakness?

9. What aspects of the service delivery strategy do you feel are most innovative? Why?

10. Which components/services do you feel have been most/least helpful for program participants?

11. a. Do you feel that the package of services offered by the program comprehensively meets the needs of participants?

b. What service gaps exist and how could each of these gaps be addressed?

c. Are there other approaches, strategies, or services that would contribute to better outcomes for program participants?

12. What is your assessment of how the responsible fatherhood program has affected the degree of service integration/coordination in the delivery of services to the target population within the community?

13. What kind of overall effects has the program had on participants, their families, and the community as a whole? To date, has the responsible fatherhood program had specific effects in any of the following areas (Note: See impact study for specific measures under each type of effect) and, if so, how:

a. Responsible Father Behavior (e.g., safe sex behavior, reduced unplanned child-bearing, marriage/stable relationships, reduced substance abuse, reduced criminal involvement, and community connectedness).

b. Father's Relationship with Child (e.g., paternity status, contact/visitation, type of child-related activities in which the father participates, parenting skills, and closeness).

c. Father's Financial Capabilities/Support (e.g., child support, employment and earnings, work ethic/attitude, education/training activities, housing, other responsibilities, physical health, mental health, self-awareness/self-esteem, anger management, and ability to deal with racism).

d. Child Well-Being (e.g., safety in the household, physical health, emotional/mental health, academic achievement, social behavior, and problem behavior).

e. Co-Parenting Relationship (e.g., arrangement for child access, agreement on child support, agreement/cooperation concerning child-rearing, parents' feelings toward each other, father's attitudes toward significant others, and quantity and quality of communication between parents).

f. Other Perceived Impacts/Effects.

14. a. To what extent has the program been able to meet the needs of the surrounding community?

b. Can you identify any specific impacts that it has had on the surrounding community?

15. Would you recommend to other agencies like your own, funding initiatives such as this responsible fatherhood program?

16. Do location, demographics, and other distinctive features at this program site make the program either non-transferable or limit its transferability?

17. What features of this responsible fatherhood program would you suggest replicating in other localities across the country? What would you suggest not replicating?

18. Do you have any suggestions to other funding agencies that might be interested in initiating a responsible fatherhood program?

Discussion Guide for Responsible Father Program Participants

A. PROGRAM CONTEXT AND INVOLVEMENT

1. Tell me about your community. What are the worst things about it? What are the best things about it?

2. How did you first hear about the <formal name for the responsible fatherhood program>?

3. Why did you decide to participate in the program?

4. Do you know eligible fathers who do not want to participate in the program? Why do you think they don't want to participate?

5. Have you participated in other programs similar to the responsible fatherhood program in the past? What is the difference between the program you are participating in and other past programs? What do you think are the differences in the types of individuals participating in this program and other past programs?

B. PROGRAM COMPONENTS AND GOALS

1. What do you believe to be the goals of the responsible fatherhood program? Do you think that these are good goals?

2. Please list all of the program services (that you know about) and indicate the ones in which you've participated.

3. Which components/services do you like the best and least?

4. Do you think that new components/services should be added to the responsible fatherhood program in the future? Which ones? Why?

5. Do you have a case manager? If yes --

a. How did you get your case manager?

b. Did you have a choice of case managers?

c. Do you like the case manager that you got? Why?

d. How often do you see your case manager? Where do you usually meet? What do you usually discuss?

e. Has your case manager ever met your family?

f. Did you get a service plan? Do you know what is in your service plan? Who was there when your service plan was developed?

g. To what kinds of services did the case manager refer you? Were the services to which you were referred appropriate?

D. PROGRAM OUTCOMES

1. How have you been helped by participating in the responsible fatherhood program? Has participation in program made a difference for you in any of the following areas?

a. Responsible Father Behavior (e.g., safe sex behavior, reduced unplanned child-bearing, marriage/stable relationships, reduced substance abuse, reduced criminal involvement, and community connectedness).

b. Father's Relationship with Child (e.g., paternity status, contact/visitation, type of child-related activities in which the father participates, parenting skills, and closeness).

c. Father's Financial Capabilities/Support (e.g., child support, employment and earnings, work ethic/attitude, education/training activities, housing, other responsibilities, physical health, mental health, self-awareness/self-esteem, anger management, and ability to deal with racism).

d. Child Well-Being (e.g., safety in the household, physical health, emotional/mental health, academic achievement, social behavior, and problem behavior).

e. Co-Parenting Relationship (e.g., arrangement for child access, agreement on child support, agreement/cooperation concerning child-rearing, parents' feelings toward each other, father's attitudes toward significant others, and quantity and quality of communication between parents).

f. Other Perceived Impacts/Effects.

2. How do you feel the program helped other participants?

3. Do you see any general improvements in your community because of the responsible fatherhood program?

4. a. Do you think that individuals such as yourself in other communities would benefit from participation in a responsible fatherhood program such as the one you have participated in? If so, how?

b. Is there anything you would change before setting the program up in another locality?

Discussion Guide for Community Leaders and Residents

1. Tell me about your community. Describe its residents. What are the strengths of the community? What are the weaknesses?

2. How long have you lived here? If you've lived here over five years, have you noticed any changes in characteristics of the population over the past five years?

3. What would you consider the top three problems within your community today? How do you think these problems should be addressed?

4. Is fatherlessness (i.e., the lack of a father within a home where children are present) a major problem within the community? If so, what are its impacts?

5. Have you ever heard of the <formal name for responsible fatherhood program>? If yes, how did you hear about the program? Do you have any ongoing involvement with the program?

[Note: Continue if the respondent has heard of the program; otherwise stop the interview.]

6. Can you identify some of the goals that the responsible fatherhood program is hoping to achieve within your community? Do you think that these are appropriate goals?

7. Do you have any general impressions about the fatherhood program and how it might be affecting your community?

8. Are there other approaches, strategies, or services that you feel are needed within your community to address the problem of fatherlessness?

Return to ToC

APPENDIX D

EVALUATION OF THE RACINE GODWILL INDUSTRIES PROGRAM

Office of Strategic Finance

Sandra Cleveland, Evaluation Section

July 17, 1996

PO Box 7850

Madison WI 53707-7850

Phone (608) 266-3816

Fax (608) 267-0358

Summary of 1994 Children First Data

This report displays evaluation data collected from the counties about Children First participants who were referred to the program in 1994. Counties routinely send data collection forms to the State describing each participants demographic characteristics, child support payment history and a follow-up form describing the services received and the extent to which participants completed the program. The data from these forms are entered into a database that will be used to prepare an evaluation report for the federal government in July 1998. That evaluation will present the outcomes of clients referred to the program in 1994, 1995 and 1996.

The 1994 database includes information for a total of 1084 cases, of which 981 cases were for regular participants in the program and 103 cases describe participants in a control group in Racine County. The following section describes the outcomes and characteristics of regular participants in the program. There remains a few missing cases which may be included before the final Children First report is produced in July, 1998.

Comparison of Child Support Payments Made by Regular Participants

The following table compares the amount of child support paid, the number of payments made and the total number of people making payments before and after referral to the program. The analysis is restricted only to those participants who had a child support order in effect for at least six months prior to referral. Cases with missing data were excluded leaving a total of 785 cases included in the analysis. Participants owed an average of $4,191.46 in delinquent child support payment at the time they were enrolled.

Table 1.

Pre-Post Comparison of Child Support Payments Made by Children First Participants

Time Period	Child Support Paid	Change	Number of Payments Made	Change	No. Paying	Percent Paying/ Percent Change
Six Months Before Referral	$205.39		2.82		288	36.7% (n/a)
First Six Months After Referral	$448.26	+$242.87 (+118.2%)	6.13	+3.31 (+117.4%)	556	70.8% (+93.1%)
Second Six Months After Referral	$445.97	+$240.58 (+117.1%)	6.11	+3.29 (+116.7%)	506	64.5% (+75.7%)

Employment Status of Program Participants

The following table describes changes in the employment status of program participants. Missing data was excluded, leaving 589 cases in the analysis. Participants had an average of 2.4 different employers in the past three years and worked approximately four months in the last year.

Table 2.

Employment Status of Children First Participants

Is the participant currently employed?	At enrollment	At case closure	Percent change
Yes	161 (27.3%)	360 (61.1%)	+199 (+123.6%)
No	428 (72.7%)	162 (27.5%)	-266 (-62.2%)
Don't Know	0 (0%)	67 (11.4%)	+67 (n/a)
Total	589 (100%)	589 (100%)	
Average Hourly Wage	$6.83	$6.38	- .45 (-6.6%)

Educational Level of Children First Participants

The following tables describe the educational level of Children First participants. Each individual analysis excludes missing data. Children First participants have an average of 11.31 years of education.

Table 3

Proportion of Participants with a High School Diploma

Does the participant have a high school diploma?	Number	Percentage
Yes	421	52.0%
No	389	48.0%
Total	810	100.0%

Table 4

Level of Training Beyond High School

Level of Training	Number	Percentage
No training	544	68.9%
Some additional training, but no diploma	149	18.9%
Associate or technical degree	79	10.0%
Bachelor's Degree	16	2.0%
Advanced Degree	2	.2%
Total	790	100%

Demographic Characteristics of Children First Participants

The following tables display some key demographic characteristics of Children First participants. A total of 981 cases were analyzed, but missing data was excluded for each analysis. The average age of Children First participants was 31.5 at the time they were enrolled in the program. Participants had an average of 2.4 children whose average age is 7.

Table 5

Race of Children First Participants

Race	Number	Percentage
White	512	59.5%
Black	260	30.3%
Hispanic	67	7.8%
Native American	21	2.4%
Total	860	100%

Table 6

Marital Status of Children First Participants

Marital Status	Number	Percentage
Single	412	48.8%
Married	130	15.4%
Widowed	2	.2%
Divorced	220	26.0%
Separated	81	9.6%
Total	845	100%

Table 7

Sex of Children First Participants

Sex	Number	Percentage
Male	863	88.4%
Female	113	11.6%
Total	976	100%

Program Completion

The following section was developed by analyzing the responses given on the Compliance Monitoring Form submitted by counties. There were a total of 882 of these forms returned for regular participants in the program for 1994 participants. Each analysis excludes missing data.

Table 8

Program Completion

Did the client complete the program within one year of enrollment?	Number	Percentage
Yes	395	63.8%
No	224	36.2%
Total	619	100.0%

Table 9

Method of Fulfilling Children First Requirements

How did the participant complete the program?	Number	Percentage
Participating in components	53	13.4%
Paying three months of child support	342	86.6%
Total	395	100.0%

Responses on the compliance monitoring form identified reasons that 103 participants were exempt from the program or that their case was closed before they completed the program.

Table 10

Exemptions and Case Closures

Reasons for exemptions and closures	Number	Percentage of all exemptions/ closures
Medically exempt (SSI)	28	27.2%
Returned to family	6	5.8%
Participant moved away and program lost contact	28	27.2%
Participant sent to prison	19	18.5%
Child support order changed/court exempted	14	13.6%
Participant gained custody	4	3.9%
Participant died	2	1.9%
Other	2	1.9%
Total exemptions and closures	103	100.0%

Table 11

Participants Sent to Jail During Participation

Was the client sent to jail during participation in the program?	Number	Percentage
Yes	106	18.1%
No	480	81.9%
Total	586	100.0%

Table 12

Reasons Participants sent to Jail

Reasons	Number	Percentage
Failed to comply with Children First or Child Support Order	44	41.5%
Other reason	37	34.9%
Don't Know	25	23.6%
Total	106	100.0%

Control Group Comparison

As part of the evaluation, a small control group was established in Racine County. Racine County was the only Children First county volunteering to develop a control group. A portion of the people identified for enrollment in Children First in Racine County were randomly assigned to the control group. This group received the same type of service they would have received if the Children First program did not exist. Prior to Children First, parents who were delinquent in their child support payments and who were unemployed were ordered by the court to conduct a job search. They were usually required to contact a minimum number of employers and report their contacts to the court. Members of the control group received these services only.

The final evaluation report of Children First will include a detailed analysis of the outcomes of the control group in order to compare their outcomes to those of regular children First participants in Racine County. This comparison will allow the evaluator to determine the extent to which the program impacts child support payments in Racine County compared to what would have occurred without the program. The following tables are some preliminary comparisons of these outcomes based on the first year of the experiment. Table 13 displays a comparison between Racine County's Regular Participants and the control group. Again, only cases where there were no missing child support data and where the child support order was in effect at least six months prior to referral to the program were included in the analysis. Table 14 displays employment outcomes for these two groups.

Table 13.

Pre-Post Comparison of Racine County's Regular Participants in Children First and Control Group

(Valid total n for Regular Participants=256. Valid total n for control group=70)

Time Period	Child Support Paid		Number of Payments Made		No. Paying/ Percent of Total	
	Regular	Control	Regular	Control	Regular	Control
Six Months Before Referral	$212.35	$256.85	2.33	2.03	85 (33.2%)	16 (22.8%)
First Six Months After Referral	$374.17 (+76.2%)*	$416.19 (+62.0%)	5.98 (+156.8%)*	6.06 (+198.5%)	176 (68.8%)	46 (65.7%)
Second Six Months After Referral	$375.06 (+76.6%)	$466.20 (+81.5%)	5.71 (+145.1%)	6.46 (+218.2%)	147 (57.4%)	38 (54.3%)

*Percentage change from pre-referral total

Table 14.

Comparison of the Employment Status of Regular Participants in Racine County's Children First Program and their Control Group (Excludes missing data)

Is the participant currently	At enrollment (number/percent of total)	At case closure (number/percent of total)	Percent change from enrollment to closure

Employed?	Regular	Control	Regular	Control	Regular	Control
Yes	54 (22.0%	10 (20.0%)	144 (58.5%)	4 (8.0%)	+90 (_166.6%)	-6 (-60%)
No	192 (78.0%)	40 (80.0%)	100 (40.7%)	37 (74.0%)	-92 (-47.9%)	-3 (-7.5%)
Don't know	0 (0%)	0 (0%)	2 (8%)	9 (18.0%)	+2 (n/a)	+9 (n/a)
Total	246	50	246	50		
Average Hourly Wage	$5.97	$5.99	$5.73	N/A	-24 (-4.0%)	N/A

Return to ToC

APPENDIX E

TECHNICAL DESCRIPTION OF PARTICIPATION AND IMPACT ANALYSES

In this appendix we present a technical description of the multivariate analyses that are likely to be required for the impact evaluation. We describe both the participation and impact analyses.

I. Participation Analysis

The details of the participation analysis will depend on which type of evaluation design is used (experimental, non-experimental, or randomized outreach) and on whether a single-site or multi-site evaluation is performed. Corresponding to the discussion in the text (Chapter Seven, Section III.C) we first describe participation analysis for an experimental, single-site evaluation, then consider modifications necessary for the alternative designs and for a multi-site evaluation.

A. Participation Analysis under an Experimental Design

Under the experimental design, we assume that only the volunteers who are assigned to the treatment group can choose whether or not to participate. The formal specification of the binomial choice model is:

Equation A.1 $P_i^* = \forall_0\_+\_\forall_1 Z_{1i} +\_\ldots\_+\_\forall_J Z_{Ji} - v_i,$

Equation A.2 $P_i = 1 \text{ if } P_i^* > 0$

$= 0 \text{ if } P_i^* < 0,$

where:

P_i^* is the value of an unobserved "participation index" for father "i" -- the probability that a father will participate is an increasing function of the participation index;

$\forall_0$ is an intercept parameter;

$Z_{1i} \ldots Z_{Ji}$ are characteristics of father i, measured through the baseline survey;

$\forall_1\_\ldots\_\forall_J$ are coefficients for the characteristics;

v_i is a random disturbance, representing other factors that might influence participation (after controlling for the Zs);[1] and

P_i is an indicator for whether the father participates in the program (P_i is one if the father participates and zero otherwise).

---

[1] The negative sign before v_i in Equation A.1 is intentional, but has no substantive implication. Given the negative sign, a positive value for $\forall_j$ means that an increase in the value of Z_j is associated with an increase in the probability of participation.

Equations A.1 and A.2 together imply:

Equation A.3: $\text{Prob}(P_i = 1) = F(\forall_0\_+\_\forall_1 Z_{1i} +\_...\_+\_\forall_J Z_{Ji})$;

where F(.) is the distribution (cumulative density) function of v_i. For the probit model, F(.) is the standard normal distribution function, and for the logit model F(.) is the logistic distribution function.

The logit model is often preferred to the probit model when a binomial choice model is estimated alone, primarily because of the simplicity of the functional form for F(.); both models usually yield very similar results when interpreted appropriately.[2] The probit model is often used when the participation equation is estimated jointly with one or more related equations. In this case it is usually assumed that the participation disturbance, v_i, is joint normally distributed with the disturbance(s) in the other equation(s). For the evaluation, we expect the participation equation to be estimated jointly with one or more outcome equations, to measure the impact of participation (see below).

The parameters of the binomial choice model would likely be estimated by the method of maximum likelihood, using data from the treatment group only. It would be inappropriate to include the control group data in this analysis because the study volunteers assigned to this group are not offered the choice of participating.

To assist in interpreting the parameters, it is helpful to calculate the marginal effect of a change in each Z on the probability of participation for the "average" father -- the father with characteristics equal to the mean characteristics for fathers in the sample. In general, the change in the probability for a change in Z_j, a typical characteristic variable, for the typical father can be calculated as:

Equation A.4: $)P_i = F(\forall_0\_+\_\forall_1 Z_{1i} +...\forall_j(Z_{j0} +)Z_j) +...+\_\forall_J Z_{Ji}) - F(\forall_0\_+\_\forall_1 Z_{1i} +...\forall_j Z_{j0} +...+\_$

$\_\_\_\forall_J Z_{Ji})$,

where Z_{j0} is the value of Z_j before the change, $)Z_j$ is the change, all other Zs are fixed at the actual values for father i. The before and after values for Z_j would depend onthe nature of the variable. For instance, the variable could be a dummy variable indicating whether a father is in one of two ategories (e.g., never married to mother vs. divorced). In the case of a dummy variable, Z_{j1} is zero and $)Z_j$ is one, so the calculated value of $)P_i$ would be the difference in the probability of participation for fathers in these two categories, holding other variables constant at the values for father i.[3] The sample mean of the $)P_i$ could be used to estimate the effect of the change on the average father.

---

[2] See Greene, W. (1990) Econometric Analysis, Chapter 20. New York: MacMillan.

[3] For continuous variables, the researcher could, instead, evaluate the derivative of F(.) with respect to Z_j at the actual value of the father's explanatory variables, then compute the mean derivative to obtain the mean effect per unit change for "small" changes.

B. Participation Analysis in a Non-Experimental Design

The appropriate methodology for a non-experimental design is the same as for the experimental design. As discussed in Chapter Seven, the situation faced by the members of the treatment group in the non-experimental design is the same as that faced by the treatment group -- referred volunteers -- in the experimental design. The parameter estimates from the analyses under these alternative designs would likely be quite different, however, simply because the treatment group is referred in to the program in one design, but not in the other.

C. Participation Analysis under a Randomized Outreach Design

In the randomized outreach design (Chapter Three), study volunteers are randomly assigned to receive strong (treatment) or weak (control) outreach. Fathers in either group may decide to participate in the program, but the differences in outreach are expected to result in higher participation rates among fathers who receive the treatment outreach.

Under this design, data for both the treatment and control groups would be used in the participation analysis because fathers in both groups choose whether or not to participate. One of the variables to be included in the Zs would be an indicator for the treatment outreach. If the evaluators use multiple types of randomized outreach, the Zs would include indicators for all types, or other variables that are used to define all of the outreach types (e.g., monetary or other continuous measures of the intensity of outreach). The coefficients of the treatment outreach variables would measure the impact of the variables on the propensity to participate, and could be easily converted to estimates of the effect of outreach on the probability of participation.

The evaluator may also want to explore interactions between the outreach variables and the other explanatory variables. This would be feasible if the sample size is sufficiently large. It may be that outreach is more effective for fathers with some characteristics than for others. For instance, a significant interaction between an outreach variable and an age group variable would suggest that the outreach is more effective for fathers in some age groups than in others.

D. Participation Analysis in a Multi-site Evaluation

The same methodology would be applied in a pooled analysis under either the experimental or non-experimental design, but the explanatory variables (Zs) need to be modified appropriately. ost importantly, dummy variables to indicate the site should be included because participation is likely to be higher in some sites than in others even after controlling for observed baseline characteristics of father fathers. The evaluator may also allow for different effects of various factors across sites. In the extreme, this could mean estimating separate models for each site, but this would result in the loss of any advantage that might be gained from pooling the data. Because sample sizes for each site are likely to be modest, it would be prudent to pool the data unless there are strong prior reasons to believe that the effects of the explanatory variables on participation vary across sites.

The participation analysis for a multi-site evaluation under a randomized outreach design should also include dummy variables to indicate the site. In addition, the evaluators may want to interact site dummies with the outreach treatment dummy or, if applicable, the multiple outreach variables. This would allow the evaluator to test the null hypothesis that the effect of the randomized outreach on participation is the same at all sites, and to estimate differences in effects across sites.

II. Impact Analysis

As with the participation analysis, we begin with consideration of the experimental design case for a single site, then discuss modifications needed for non-experimental, random outreach, and multi-site designs. We also begin with the assumption that the outcome variable is continuous and has an unlimited range, then consider modifications needed for categorical and limited dependent variables.

A. A Model for a Continuous Outcome Variable under an Experimental Design

The econometric model we describe below is a standard model for the impact of a randomized treatment on a continuous outcome variable with an unlimited range.[4] We begin with a model that is appropriate for an experiment in which all treatment group subjects actually receive the treatment -- in this case participate in the fatherhood program. Under the experimental design described in Chapter Three, however, some treatment group members who are referred to the program chose to not participate. Hence, we modify the model in an appropriate way.

The model relates the outcome variable to whether or not the subjectparticipates in the program, characteristics of the subject observed at baseline, and unobserved, random factors;

Equation A.5: $Y_i = *P_i + \exists_0\_+\_\exists_1 X_{1i} + ... + \_\exists_K X_{Ki} + ,_i,$

where:

Y_i is the value of the outcome variable for father i, measured in the follow-up survey (high values are associated with responsible behavior);

P_i is a dummy variable indicating whether father i was a program participant, as previously defined (1 = participant, 0 = non-participant);

$*$ (delta) is the coefficient of P, to be estimated;

X_{1i} ... X_{KI} are baseline characteristics for father i;

$\exists_0$ is the equation intercept, to be estimated;

$\exists_1$ ..._$\exists_K$ are coefficients for the baseline characteristics, to be estimated; and

$,_i$ is a random "disturbance" -- factors that affect outcomes and are assumed to be independent of the father's baseline characteristics, to be independent across fathers, and to have constant variance across independents.

If all subjects who are randomly assigned to the treatment group participate in the program, and all control group subjects do not, then the participation dummy, P, is synonymous with a dummy for assignment to the treatment group. Further, under the same condition the dummy would be independent of the disturbances in the equation because of random assignment. Under such circumstances, the unbiased estimates of the parameters of the regression model can be obtained by the method commonly referred to as "ordinary" least squares (OLS); the OLS coefficient of P (i.e., the estimate of *) will be an unbiased estimate of the impact of the program on Y.

If, instead, some randomly assigned treatment group subjects elect not to participate, along with all control group subjects, then P will have a value of zero for some treatment group subjects (i.e., the non-participants). Further, among treatment subjects the value of P is likely to be correlated with the random disturbance, ,, violating a key assumption of the regression model.

---

[4] See Maddala, op cit.

*The OLS estimate of * would then be a biased estimate of the impact of participation.*

To understand the source of bias, consider two fathers with identical characteristics (Xs) but, for reasons that are unobserved, one father is more highly motivated to behave responsibly toward his child than the other. The more highly motivated father is both more likely to participate in the program ($P = 1$) and more likely to have a better outcome (high value for Y) holding participation constant (high value for ,). Under these circumstances, a least squares estimate of * would overstate the impact of the program (i.e., be biased upward) because it would capture the factthat treatment group members who choose to participate are typically more highly motivated than those who don't, holding observed baseline characteristics constant.

The findings from the participation analysis can be used to eliminate the bias, as follows. For treatment subjects, define $u_i = P_i - F_i$, where $F_i = F(\forall_0\_+\_\forall_1 Z_{1i} +\_...\_+\_\forall_J Z_{Ji})$ is the probability that a treatment subject with baseline characteristics Z_{1i} ... Z_{Ji} in the program, as defined above. The variable u_i is a random disturbance -- the random difference between the indicator of participation and the probability of participation -- and is independent of the values for the Zs. For control group subjects, for whom P_i is always zero, u_i and F_i are always zero, by definition.

Turning the definition of u_i around gives:

Equation A.6: $P_i = F_i + u_i$.

Substitution of the right-hand side of Equation A.6 for P_i in Equation A.5 yields:

Equation A.7: $Y_i = *F_i + \exists_0\_+\_\exists_1 X_{1i} +...+\_\exists_K X_{Ki} + *u_i\_+\_,_i,\_$

$\_\_\_\_=\_*F_i + \exists_0\_+\_\exists_1 X_{1i} +...+\_\exists_K X_{Ki} + ,_i*,$

where $,_i* = *u_i\_+\_,_i$. The transformed disturbance, $,_i*$, is independent of F_i by definition, and will also be independent of the Xs if all of the Xs are included in the Zs, as seems likely to be appropriate. Were F_i observed, unbiased estimates of the coefficients of this model, including *, could be obtained by OLS; i.e., by regressing Y on F (including zeros for the control group cases) and X. More efficient estimates could be obtained by weighted least squares (WLS), taking into account the fact that variance of the transformed disturbance varies across subjects.[5]

Although F_i is not observed, OLS or WLS can be successfully applied to the same model after substituting estimated values of F_i for the treatment subjects, obtained from the participation analysis. The estimated treatment effect will be unbiased if the sample size is reasonably large. Estimated standard errors need to be adjusted to take account of the fact that estimated, rather than actual, values for F are used.[6] One alternative estimation method applies OLS to Equation A.5, then adjusts the estimated coefficient for bias due to non-participation.[7] A third alternative jointly estimates the parameters of the participation and outcome equations via maximum likelihood or some other joint method.[8]

---

[5] See Maddala, op cit.

[6] See Maddala, op cit.

[7] See Bloom, op cit.

[8] Maximum likelihood estimation requires more restrictive assumptions on the distribution of $,_i$ than we have made here. The most common assumption made is that $,_i$ and v_i, the disturbance in the participation model, have a joint normal distribution. See Maddala, op cit.

Two features of this methodology deserve further attention before we turn to variants for alternative evaluation designs. First, the method can be used to estimate participation effects even if there is no control group other than self-selected non-participants, but is not likely to work well. In such a case, it would be essential that some elements of the characteristics that determine participation, Zs, not be included in the Xs. If not, the estimated values of F will be highly (multi-) collinear with the Xs, resulting in a very imprecise estimate of the program impact; exact collinearity would be avoided only by the fact that F is a non-linear, rather than linear, function of the Zs. Strong candidates for variables to include in the Zs, but not the Xs -- variables that have a strong effect on the probability of participation but only a negligible direct effect on the outcome variable -- are hard to find. In this case it would also be extremely important to include controls for possible systematic differences between the participant and non-participant groups in the outcome equation.

Inclusion of variables in the Zs that are not in the Xs will be helpful in improving the recision of estimates when there is a control group, too, but such variables are not critical in this case because the zero values of F for the control group will eliminate the high collinearity that would exist between F and the Xs if only treatment subjects were included in the analysis. We will return to this issue in the discussion of the methodology for a randomized outreach evaluation, where it is more critical.

A second feature of this methodology is that it assumes that program participation has the same impact for all participating fathers. This seems unlikely. A much more general model would specify entirely different relationships between the outcome variable and baseline characteristics (Xs) for participants and non-participants; i.e., participation would be modeled as changing the entire relationship between the baseline characteristics and the outcome variable, rather than a "parallel shift" of the equation.[9] Under this model the impact of program participation would vary with baseline characteristics in a very nonrestrictive way.

The sample sizes that would be required to obtain reasonably precise estimates of such a general model are not likely to be achieved given the size of current responsible fatherhood programs. We recommend, instead, that the assessment of variation in impacts with baseline characteristics be limited to examining interactions between impacts and a very small number of key characteristics, assuming that the effects of other baseline characteristics on outcomes are invariant to participation. For instance, the evaluator might investigate the relationship between the size of the impact and a baseline characteristic variable W_i, say, using the specification:

Equation A.8: $Y_i = *_{\equiv}\_P_i + *_{w}\_(P_i W_i) + \exists_0\_+\_\exists_1 X_{1i} +\ldots+\_\exists_K X_{Ki} + ,_i,$

in which the impact of the program for a father with characteristic W_i is $*_{\equiv}\_+ *_{w}\_W_i$. If W were a dummy variable, classifying fathers into one of two groups by a baseline characteristic, then $*_{\equiv}$ would be the impact for the base (zero) group and $*_{w}$ would be the difference between the impacts for the two groups.

B. Application to a Non-Experimental Design

As discussed in the text, the methods described for the experimental design can also be used for the non-experimental design. The main difference has to do with the importance of the baseline characteristics (Xs and Zs). They are more important inthe non-experimental design because distributions of these variable in the treatment and control groups may differ substantially.

---

[9] See Maddala, op cit.

C. Modifications for a Randomized Outreach Design

In the randomized outreach design (Chapter Three), study volunteers are randomly assigned to receive strong (treatment) or weak (control) outreach. Fathers in either group may decide to articipate in the program, but the differences in outreach are expected to result in higher participation rates among fathers who receive the treatment outreach.

The same methodology can be applied to estimating the impact of participation on an outcome after making these modifications. First, measures of the randomized outreach treatment should be included in Z. In the simplest case, this would be a dummy variable to indicate whether the father received the treatment or control outreach. As discussed in the previous chapter, however, the evaluators may want to try several variants of the treatment outreach, so more than one variable may be needed to capture the outreach treatment.

Second, the participation analysis will be performed using both treatment and control group cases, and the probability of participation, F_i, will be positive for control group cases as well as for treatment group cases; recall that this probability was zero by definition in the experimental and non-experimental models.

In all other respects the model for the experimental design applies. With the modifications in place, unbiased estimates of the participation effect, *, can be obtained by the same means as would be applied in the other cases.[10]

The role and importance of effective treatment outreach becomes evident by recognizing that this model is formally equivalent to a model discussed in Section II.A, above, in which all volunteers are self-selected into participant or non-participant groups. We criticized that model on the grounds that the Zs were likely to include the same variables as the Xs and, as a result, there would be high collinearity between F and the Xs. The randomized outreach serves to break up this collinearity; it would presumably only affect outcomes through its effect on participation, and would not be included in the Xs.

The role of randomized outreach in the estimation methodology implies that the outreach must satisfy two important criteria. First, it must be effective; if it does not have a substantial impact on the probability of participation it will do little to reduce the collinearity between F and the Xs. Second, it should have a negligible direct effect on outcomes. Some outreach methods might have substantial direct effects: significant moral suasion from a respected role model, or promises of long-term financial or other material rewards for participating are examples. Such methods might also be very effective in increasing participation, so some care mustbe exercised to avoid them.

D. Extension to Categorical and Limited Dependent Variables

To this point we have assumed that the outcome (dependent) variable is a continuous variable with unlimited range. It is likely, however, that many key outcome variables will not satisfy both of these conditions. Some will be categorical (e.g. paternity establishment) while others will have a limited range (e.g., hours of child contact and level of child support cannot be negative). Further, among categorical variables there are likely to be two types: qualitative variables, that indicate which f two unranked categories a father is in, and ordinal variables, where the categories have a meaningful ranking from lowest to highest (e.g., responses to questions that require selection of a value from a numerical scale).

Appropriate modifications to the regression model can be made to accommodate each of these types of dependent variables. Possibilities include:

---

[10] See the discussion following Equations A.6 and A.7.

Probit and logit for binomial dependent variables (qualitative or ordinal);

Multinomial probit and logit for multinomial (more than two categories) qualitative dependent variables;

Ordered probit for ordinal multinomial variables; and

Tobit and many other limited dependent variable models for dependent variables with a limited range.[1]

Each of these models can be characterized in the following general way:

Equation A.9: $Y_i^* = *P_i + \exists_0\_+\_\exists_1 X_{1i} +...+\_\exists_K X_{Ki} + ,_i,$

Equation A.10: $Y_i = g(Y_i^*; 2),$

where:

Y_i^* is an unobserved index variable that is continuous and has unlimited range;

$g(Y_i^*;2)$ is a parameterized function that maps the index variable into the observed dependent variable; and

2 is a set of parameters for g(), to be estimated along with other parameters of the model.

All other notation is as defined previously.

Equation A.9 is identical to Equation A.5 except that the "dependent variable," Y_i^*, is not observed directly. Equation A.10 specifies the relationship between the dependent variable that we do observe and the unobserved variable. This equation is deterministic.

---

[1] See Maddala, op cit.

The probit and logit models provide very simple examples of Equation A.10. Both specify that Y_i is zero or one, depending on whether Y_i* is below or above zero.[1] In this case g() does not have any parameters. The ordered probit model is an example in which g() has parameters. In this case,

quation A.11: $Y_i = 0$ if $Y_i* < 2_1$;
$= 1$ if $2_1 < Y_i* < 2_2$;
$= 2$ if $2_2 < Y_i* < 2_3$;
.
.
.
$= M$ if $2_M < Y_i*$;

where M is the largest value of Y_i. The parameters (2) are, in these case, threshold values of the index variable.

In the absence of a selection issue for P_i, models such as these can be estimated using appropriate single-equation methods. Most frequently, the maximum likelihood approach is applied. If selection is an issue, the preferred approach is likely to be joint estimation of the outcome and participation equations, by maximum likelihood or perhaps by some method that is less computationally intensive. Joint estimation usually requires specification of a joint distribution for the disturbance in the outcome equation, $,_i$, and the disturbance in the index equation for the participation model, v_i (see Equation A.1). The most commonly used assumption in such situations is that the two disturbances have a joint normal distribution.

E. An Econometric Model for Jointly Analyzing the Impacts of Multiple Programs

In this section we begin by modifying the methodology discussed above for the estimation of impacts in an experimental design for the evaluation of one program to the joint evaluation of multiple programs (including multiple sites for a single program). We assume that volunteers at each site are randomly assigned to control and treatment groups, that some treatment subjects do not participate in the program at each site, and that all control subjects do not participate. We also assume there is no cross-site contamination (e.g., subjects at one site participating in the program at another site.) We then turn to using the modified model in non-experimental and randomized outreach designs.

1. Experimental Design

The following modification of the single-site model for a randomized design can be used to jointly evaluate the impacts of the program at all sites:

Equation A.12: Y_i = $*_1P_{1i}$ + $*_2P_{2i}$ + ... + $*_mP_{Mi}$ + $(_1S_{1i}$ + $(_2S_{2i}$_ + ... + $(_mS_{Mi}$_+_$\exists_0$_+____$\exists_1X_{1i}$ +...+_$\exists_KX_{Ki}$ + $,_i$, where:

---

[1] The models are distinguished by the distributional assumption for ,.

$_{1i}$, P_{2i}, ... P_{Mi} are dummy variables to indicate participation at sites 1 through M (P_{mi} is one if the subject participated at site m, and is zero otherwise);

$*_1$, $*_2$, ... $*_m$ are coefficients of the participation dummies, and represent the impacts of the respective programs on the outcome variable;

S_{1i}, ... S_{Mi} are dummy variables to indicate which site the subject is from, regardless of participation (S_{mi} is one if the subject is a volunteer at site m, and is zero otherwise); and

$(_1$ ... $(_m$ (gammas) are coefficients of the site dummies, and represent effects of unobserved factors that vary across sites ("site effects").

All other notation is as defined previously. The variance of the disturbance term, ,, could be assumed to be constant within sites, but vary across sites.

If all subjects assigned to the treatment group in each site actually participate in the site's program, unbiased estimates of the impacts of each program can be obtained by OLS, or by WLS (in recognition of cross-site differences in the variance of ,). In the more likely case of partial participation by treatment subjects at each site, it will be necessary to replace the participation dummies in Equation A.12 with participation probabilities. These probabilities would be estimated in the participation analysis, which could be performed site-by-site, but would likely be performed jointly for all sites. Alternatively, as in the single-site methodology, the outcome and participation models could be estimated jointly.

The site dummies in Equation A.12 deserve further comment. We anticipate that outcomes will vary across sites in the absence of the programs and holding constant observed baseline characteristics of volunteers because of environmental factors. The site dummies capture the average effect of this variation on outcomes at the respective sites. They can be viewed as additional elements of X.

This methodology is well suited for comparing estimated impacts across sites. For any pair of sites, the difference in impacts can be estimated as the difference between the corresponding estimates of $*$, and a t-test for the null hypothesis of "no difference" can be easily performed. If the difference is not statistically significant, the evaluator may improve the precision of the estimates by constraining the estimated impacts for the pair of sites to be the same. This would be especially appealing for programs that are similar with respect to key program characteristics(e.g., multiple sites of a single program).

This model, like the model for a single site, assumes that participation effects are constant for all participants within a site. It also assumes that effects of baseline characteristics are the same on all subjects, regardless of site. A more general model would relax both of these assumptions. We anticipate, however, that sample sizes will be too small to obtain meaningful estimates of such an unrestricted model. We would recommend selectively relaxing these assumptions through interaction effects, analogous to the example provided for the single-site model.

B. Non-Experimental Design

As in the single-site case, the methodology developed for the experimental design can be reasonably applied to the non-experimental design if careful attention is paid to measuring baseline characteristics that are predictive of outcomes. We assume that there would be a comparison group for each site and that each comparison group site would be matched to its corresponding treatment site on environmental characteristics that are likely to have an impact on outcomes. Under this condition, the site dummies in the model would capture the environmental factors common to each site.

An alternative would be to have a different, perhaps smaller, number of comparison sites than treatment sites. In the absence of matches for each site, the site dummies would have to be dropped. They could be replaced with a set of variables that measure key aspects of the environment at each site, including the treatment sites (e.g., strength of the local labor market). The number of such variables would have to be small relative to the number of sites to obtain meaningful results.

C. Randomized Outreach Design

Under the randomized outreach design the specification for the outcome equation would be the same as under the experimental design. In this case it is clear that the participation dummy variables cannot be treated as exogenous variables. For both treatment and control groups, these dummies need to be replaced by participation probabilities, from the participation model. The participation model itself would use data from all sites and its explanatory variables (Zs) would include both site and treatment dummies.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)

PS030238